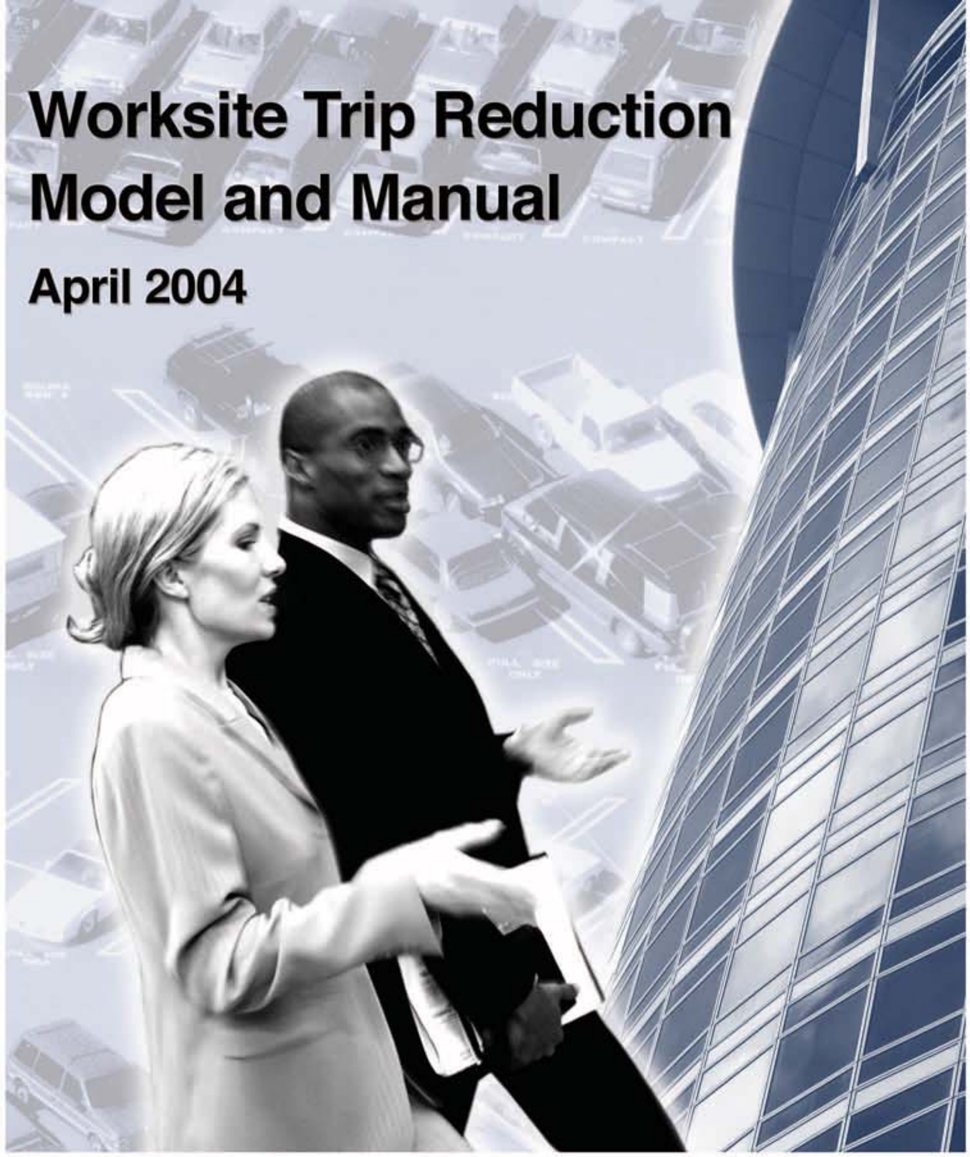
Worksite Trip Reduction Model and Manual

April 2004



WORKSITE TRIP REDUCTION MODEL AND MANUAL

April 2004

Prepared for

Florida Department of Transportation
BC-137-41

Prepared by

National Center for Transit Research
Center for Urban Transportation Research
University of South Florida
4202 E. Fowler Avenue, CUT100
Tampa, FL 33620-5375
(813) 974-3120
http://www.nctr.usf.edu/

Florida Department of Transportation Project Manager:
Michael Wright, Commuter Assistance Program Manager

Principal Investigators:
Philip L. Winters, TDM Program Director, CUTR
Rafael A. Perez, Computer Science & Engineering, USF

Project Staff:
Jennifer Perone, Research Associate, CUTR
Ajay Joshi, Graduate Research Assistant, Computer Science & Engineering
Aaron Gregory, Graduate Research Assistant, Computer Science & Engineering
Jonathan Hershberger, Grad. Research Assistant, Computer Science & Engineering

Prepared in cooperation with the State of Florida Department of Transportation, or the US Department of Transportation

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NCTR-473-14, FDOTBC137 RPWO#41		

4. Title and Subtitle	5. Report Date
Worksite Trip Reduction Model and Manual	April 2004
	6. Performing Organization Code
7. Author(s) Winters, Philip L.; Perez, Rafael A.; Perone, Jennifer S.; & Joshi, Ajay.	8. Performing Organization Report No.
9. Performing Organization Name and Address Center for Urban Transportation Research 4202 E. Fowler Avenue, CUT 100 Tampa, FL 33620-57350	10. Work Unit No. 11. Contract or Grant No. DTRS98-G-00329
12. Sponsoring Agency Name and Address Office of Research and Special Programs U.S. Department of Transportation Washington, DC 20590 Florida DOT 605 Suwannee Tallahassee, Florida 32399	13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a Grant from the USDOT Research and Special Programs Administration, and the Florida Department of Transportation.

16. Abstract

According to Institute of Transportation Engineers, assessing the trip reduction claims from transportation demand management (TDM) programs is an issue for estimating future traffic volumes from trip generation data. To help assess those claims, a Worksite Trip Reduction Model and Manual was produced using existing data on programs, services and incentives contained in thousands of before and after worksite trip reduction plans. Models were built using linear regression and non-linear neural networks with the change in vehicle trip rate (VTR) as the dependent variable. No single variable selection technique, data handling method, or modeling approach yielded the best-fitting model for all three urban areas. The neural network model built on equally sampled data was the best generalized model based on three performance measures: the accuracy across the moderate range of change in VTR; the accuracy on full range of change in VTR and the R-square between the actual change in VTR and the predicted change in VTR. Worksite trip reduction plans explain a modest portion of the change in vehicle trip rates from one year to the next. The smaller datasets may have affected neural networks ability to identify correct non-linear relationship by overfitting the training data and reducing the neural network model's power to generalize over unseen validation data. The aggregate nature of the data loses the ability to explain whether the change in mode behavior was influenced by the programs or changes in the workforce or other exogenous variables. Quality control issues with the provided datasets affected the model building process. Future research should examine organizational culture, management styles, the total expenses incurred by the employer, employee demographics and changes in the local economy. Finally, this Worksite Trip Reduction Model and Manual should be allowed to evolve like the ITE's Trip Generation Manual, which is in its sixth edition. Efforts to improve, maintain, and disseminate such a document are critical to its widespread application and improved understanding about transportation demand management program effects on traffic volumes and parking demands.

17. Key Words	18. Distribution Statement
Transportation Demand Management, TDM, trip reduction, Carpool, Vanpool Telecommuting, Transit, Compressed Work Weeks, Flextime, Neural Network, Trip Generation	Available to the public through the National Technical Information Service (NTIS), 5285 Port Royal Road, Springfield, VA 22161, (703) 487-4650, http://www.ntis.gov/, and through the NCTR web site at http://www.nctr.usf.edu/.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	139	

Form DOT F 1700.7 (8-69)

Acknowledgements

The research team acknowledges the cooperation and support of many individuals, including FDOT project manager, Michael Wright. In addition, we would like to extend our sincere thanks to Ed Hillsman and Brian Lagerberg with the Washington State Department of Transportation, Rita Hildebrand with the Pima Association of Governments, and Antonio Thomas and Carol Gomez with the South Coast Air Quality Management District for providing the data and helping explain the nuances.

TABLE OF CONTENTS

List of Tables

EXECUTIVE SUMMARY

Today's transportation professionals and others often use Institute of Transportation Engineers' (ITE) Trip Generation Manual and the Parking Generation Manual for estimating future traffic volumes to base off-site transportation improvements or identifying parking requirements. Planners may use these same resources to evaluate the impacts of land use or zoning changes on the transportation system. According to ITE, one of the many issues facing users of this trip generation data is assessing the claims that "specific transportation demand management programs and transit services will reduce site trip generation by a certain amount." This research project was undertaken to help transportation planners and engineers assess those claims.

Using literally thousands of before/after trip reduction plans from employers with 100 or more workers, a variety of analytical techniques were applied to predict the change in vehicle trips due to worksite demand management programs and policies. The products of the effort are this Worksite Trip Reduction Manual and web-based model.

The Worksite Trip Reduction Model (WTRM) (http://www.nctr.usf.edu/worksite/) predicts the extent that each incentive, disincentive, or program would impact traffic volumes and parking needs in a specific worksite. This model would allow transportation engineers, local planners, developers, employers, and transportation demand management professionals to easily input various programs, incentives, disincentives, and worksite characteristics to obtain predictions about the change in vehicle trips from that mix of tactics.

The development of this manual and model will also save those stakeholders time and money. These tools will allow for a quick assessment of different worksite-based transportation demand management strategies on traffic volumes and parking impacts. WTRM could be used to assess parking needs of new developments, thereby reducing the cost of parking construction. Employers and developers would also know what types of services and programs to offer employees, residents and tenants to decrease onsite traffic congestion and reduce their need for parking. And, finally, reduction in vehicle traffic and vehicle miles traveled will improve air quality in the region.

Methodology

This project used several thousand worksite trip reduction plans to build the model. The data came from three urban areas in the United States: Los Angeles, Tucson, and Washington State that have had trip reduction requirements on employers for many years. Employers were required to submit plans to reach a particular objective such as a reduction in the levels of single occupant vehicle (SOV) use. The data consisted of worksite modal characteristics aggregated at the employer level and a listing of incentives and amenities offered by employers. The Los Angeles data contained the largest data set data, with the Tucson, Arizona and Washington state data sets being considerably smaller. Data quality control problems reduced the size of the data in each area and eliminated or restricted some potentially useful variables (e.g., dollar values of some incentives). For performance evaluation the datasets were divided in two disjoint sets 'training/testing set' which was used to build the models and 'validation set' which was used as an unseen data to evaluated the models.

The dependent variable chosen was the change in vehicle trip rate (VTR) (e.g., reduction of 4.5 vehicles per 100 employees). VTR correlates closely with the goals of TDM -- reduce trips, decrease air pollution, decrease the need for parking -- and is generally proportional to the desired result. Alternative dependent variables such as SOV share or average vehicle ridership (AVR) have disadvantages. SOV share misses the benefits of moving from one non-SOV mode to another where the switch may actual reducing traffic but not affect SOV share (e.g., carpool to transit). The reduction in vehicle trips is distorted when using AVR as the dependent variable due to the non-liner relationship between AVR and vehicle trips. For example, increasing AVR by 0.25 from 1.10 to 1.35 persons per vehicle would require a reduction in 17 vehicle trips per 100 people. The same increase (0.25) for a worksite with an AVR of 1.50 to 1.75 would only require a reduction in 9 vehicle trips per 100 people.

Two approaches were used for the model building process: linear statistical regression models and non-linear neural networks. The linear statistical regression models were used as a benchmark for the validity and accuracy of the neural net models. The linear statistical regression models minimize the sum of the error between the real and predicted data, learning simple linear relationships between the worksite characteristics, incentives and the dependant variable 'change in VTR', while the neural networks learn more complex non-linear relationships. Sometimes linear regression methods were used to determine which variables the neural net would use to build its models.

Several phases were followed to build the models. Models were built for each of the three datasets using a variety of approaches of handling the data, including variable selection, grouping of incentives, and the treatment of outliers. Models were also built after combining the data from the three urban areas into a single dataset. Under the assumption the transportation industry was most interested in a model that predicted when large reductions could be achieved, the model performance objective was focused on predicting the change equally well across the range of the changes in VTR.

Results and Conclusions

No single variable selection technique, data handling method, or modeling approach yielded the best-fitting model for all urban areas. In many cases, there was no significant performance difference between the top models, so the recommended model for some dataset had to be decided by using the F-value measure, which incorporated two other metrics: *Recall* which gave a measure of the completeness of the model, and *Precision* which gave a measure of the correctness of the model.

The best model for each city also was not the model that used data only from that city. Before combining the Los Angeles data set with those from the other two areas, the preferred model was the one built on the grouped incentives data with records with 'no incentives' removed. But after combining the datasets, the neural network model built with no variable selection performed better for Los Angeles than the model built with only data from Los Angeles. Also for Tucson data, a neural network model built on the equally sampled (i.e., each dataset contributes equally) data performed better than the previously selected neural network model built on the full sample (i.e., all valid records from the dataset) with ungrouped incentives data. The best model for the Washington data was the linear forced enter regression model built on full sample with ungrouped incentive data.

The generalized models for any urban area were built on the combined training datasets and equally sampled training datasets. The models built using equally sampled datasets were the ones which were not biased towards the any dataset. Overall, the best generalized model for any location is the neural net model built on equally sampled data based on three performance measures. The first performance measure is the accuracy across the moderate range of change in VTR. The second performance measure is the accuracy on full range of change in VTR. The third measure is the R-square between the actual Delta_VTR and the Predicted Delta_VTR. This is the model at http://www.nctr.usf.edu/worksite.

Overall, the neural net models performed better than the linear regression models. This might be due to the ability of the neural network program to move beyond simple linear regression, which tries to minimize the error between the predicted and actual data sets. The neural network models in many situations were able to learn the non-linear relationships among various combinations of strategies. There were some neural network models which performed worse than the linear models. This might be due to the over-fitting of the training data and reducing the neural net's power to generalize over unseen validation data.

Quality control issues with the provided datasets affected the model building process. In the case of the Los Angeles data, there were many worksites for which some incentives were available in one year, then not a single incentive was shown in the following year, only to have incentives "reappear" the next plan cycle. These unexplained gaps in reporting can affect the ability of the model to estimate the impact of a particular incentive.

The quality of the financial data and the limited information on the levels of financial support provided by employers for individual employees hindered the use of such data that may have helped explain more of the changes. Not every community asked for the dollar value of transit subsidies, for example.

Another data problem encountered was the use of different units of measurement across programs. For example, the transit subsidy values were reported ranging from $0.20 to $3,000. Though employers were to report these values as "cost per employee per month", the lower figure basis appears to be along the lines of "cost per employee per trip" and the upper figure might be the "total cost per employee using the mode per year". The difficulty is there is no way of telling from the data. Also it was found that the incentives had differences in their definition across different datasets which introduced error into the results. This problem made model-building a complex task when trying to condense and collapse all of the similar variables into one.

The aggregate nature of the data loses the ability to explain whether the change in mode behavior was influenced by the programs or changes in the workforce or other exogenous variables. While hundreds of thousands of employee data, including employee's preferences to particular options, were available from the State of Washington dataset, there was no identifier to track individual changes in behavior over time. The other two datasets did not have any comparable data to Washington's individual survey responses. Access to such disaggregate data could improve the ability of a model to track behavior changes over time based on changes to worksite incentives, amenities and programs.

FUTURE WORK

Given the common interest shared by the public sector and worksites in assessing the relative effectiveness of worksite trip reduction program, future work should begin by improving the quality of the data already being collected. Quality can be improved by adopting standard definitions and common terminology to make more use of the data collected by others to improve the accuracy of the model under different circumstances. Common terms will contribute to an expanded dataset by making the data compatible with other data from other parts of the country. This approach could be facilitated by the creation of a centralized database.

Adhering to quality control procedures also could add more explanatory power to the data, especially as it relates to the value of financial incentives. Improving the quality of information that already exists can help worksites more cost-effectively deliver vehicle trip reduction programs.

The data used in this model-building approach was aggregated to the worksite level prior by the employer. This aggregate level detail does not allow for analysis to determine individual changes in travel behavior. In order to get a real handle on what makes VTR increase or decrease and what causes people to choose alternative transportation options, attempts to control for the differences should be used. Access to disaggregate data collected over time can help establish a "test" and "control" group approach to control for differences, for example, in the composition of the workforce.

Access to the disaggregate data also would help track the long-term effects of the programs. The current project assesses the impacts between two time periods (usually separated by only one to two years). However, the cumulative effect of these programs over time is less understood (i.e., will the worksite experience a constant, variable, or exponential change in VTR over time as the programs diffuse within the workforce and move beyond the "early adopters"?) *Diffusion* is the process by which the trip reduction program offerings are communicated through certain channels over time among the employees at that location. While the collection of individual data may be difficult due to privacy and attrition issues, it is worth investigating the possibility of collecting this type of transportation behavior data to help develop sustainable transportation strategies and programs for the future.

Worksite trip reduction plans explain a modest portion of the change in vehicle trip rates from one year to the next. Future research should build on this research by examining other factors that could explain more of the variance. Organizational culture and management styles are two factors hypothesized as having a significant effect on performance. These factors may affect management support at all levels, including the selection and support for the organization's employee transportation coordinator (ETC). ETCs may have the responsibility but not the authority or resources to carry out an effective program. Turnover at the ETC position can also affect continuity of service delivery and the quality of the effort. Other factors such as the total expenses incurred by the employer, employee demographics and changes in the local economy also should be examined.

Finally, this Worksite Trip Reduction Model and Manual should be allowed to evolve like the ITE's Trip Generation Manual, which is in its sixth edition. Modest levels of financial and technical assistance efforts will be needed to make this happen. Efforts to improve, maintain, and disseminate such a document are critical to its widespread application and improved understanding about transportation demand management program effects on traffic volumes and parking demands.

I. Introduction

For years, transportation engineers and urban planners have relied on the Institute of Transportation Engineers' Trip Generation Manual and the Parking Generation Manual to estimate the traffic impact and parking needs of new developments. The trip generation estimates predict the traffic volumes of these development proposals and decisions regarding whether to allow new developments are tied to these estimates. The link between development and available capacity is especially important in areas with insufficient roadway "supply" to absorb the new development "demand" without degrading the level of service.

Both the ITE trip generation manual and the Parking Generation Manual are missing modifications to the demand side of the equation. According to ITE, one of the many issues facing users of this trip generation data is assessing the claims that "specific transportation demand management programs and transit services will reduce site trip generation by a certain amount."

This research project was undertaken to help transportation planners and engineers assess those claims. Using literally thousands of before/after trip reduction plans from employers with 100 or more workers, a variety of analytical techniques were applied to predict the change in vehicle trips due to worksite demand management programs and policies. The products of the effort are this Worksite Trip Reduction Manual and web-based model.

This project seeks to help these transportation professionals by compiling several thousand worksite trip reduction plans that have been developed and tracked for several years from three urban areas in the United States: Los Angeles, Tucson, and Washington State. These data sets were compiled into models of best fit via both Neural Net and regression model building with statistical software.

The model, VTR, predicts the extent that each incentive, disincentive, or program would impact traffic volumes and parking needs in a specific worksite. This model would allow transportation engineers, local planners, developers, employers, and transportation management associations to easily input various programs, incentives, and disincentives to obtain predictions about the change in vehicle trips from that mix of tactics.

Chapter 2 of this report summarizes the data from each of the three areas. Chapter 3 provides a description of the model building approaches. Chapter 4 provides the results of the models for each of the three areas, including the best model for each area as well as the model built with the combined and sampled datasets. Chapter 5 provides a series of lookup tables to complement the web-based model, VTR. Chapter 6 provides future research ideas. Chapter 7 provides a summary and Chapter 8 provides conclusions.

OBJECTIVE

To develop a Worksite Trip Reduction Model and a Manual that will estimate the impacts of various combinations of transportation demand management (TDM) strategies in reducing vehicle trips.

II. DATA DESCRIPTION

The data used in this project was obtained from three different locations: Los Angeles, California; Tucson, Arizona; and several locations in the state of Washington. The following sections describe the available variables for each data set.

LOS ANGELES DATA

The Los Angeles data was obtained from South Coast Air Quality Management District for Los Angeles and consisted of 33,092 total records from 7,626 company worksites. Each record represents information from a specific company worksite for a specific year. There can be several records from a specific worksite as well as several records for a company if that company has multiple worksites. The information in a record includes worksite characteristics such as shares of different modes of transportation used by employees for commuting, the different incentives used by the company to entice employees to use other commuting modes different from driving alone, etc. for a specific company. The names of incentive plans in the dataset followed a coding convention, so that the names with same first two letters were closely related. This knowledge was used in some cases to group several individual incentives into one. The reason for this was to avoid collinearity of the variables, or shared variance. By grouping similar variables together, it cuts down on this problem of shared variance and produces fewer variables with which less complex models can be built. These grouped variables also give more explanatory power to the model-building process.

The incentive grouping and field descriptions for Los Angeles are shown in Table 1. Each record included a total of 95 fields. Several years of data was collected from 1988 to 2001. Some worksites collected data only for a subset of these years. The vehicle trip rate for each worksite was calculated from the given data using -

VTR = 100*(CAR1 + MOTORCYCLE + CAR2/2 + CAR3/3 + CAR4/4 + CAR5/5 + CAR6/6 + VAN_CUTR/7) / (CAR1 + MOTORCYCLE + CAR2 + CAR3 + CAR4 + CAR5 + CAR6 + VAN_CUTR + BUS + TRANSIT + WALK + BIKE + TELECOMMUTE + CWW336 + CWW440 + CWW980)

Where,
CAR1 – Number of employees driving alone
MOTORCYCLE – Number of employees commuting by motorcycle
CAR2 – Number of employees commuting two together
CAR3 – Number of employees commuting three together
CAR4 – Number of employees commuting four together
CAR5 – Number of employees commuting five together
CAR6 – Number of employees commuting six together

VAN_CUTR – Number of employees commuting in van
BUS – Number of employees commuting by bus
TRANSIT – Number of employees commuting using transit
WALK – Number of employees commuting walking
BIKE – Number of employees commuting by bike
TELECOMMUTE – Number of employees telecommuting
CWW336 – Number of 3/36 days off
CWW440 – Number of 4/40 days off
CWW980 – Number of 9/80 days off

Given that the objective was to build a model that could predict the effect on VTR of one or more incentives introduced by a worksite, the data described above was used to calculate the change in VTR (96^{th} field) between consecutive years. For example, worksite 'A' in 1999 had a VTR of 90 and in 2000 had a VTR of 85, then the difference in VTR (85 – 90) of –5 was associated with the 1999 record for that worksite. All the last year records for all the worksite were deleted, as change in VTR for them cannot be calculated. This associated data set now consisting of 25459 records (full sample data) was used to build the models described in this report.

Table 1: Los Angeles Data variable description and grouping

No.	Field Name	Description	Grouping
1	Motorcycle	Share of employees commuting by motorcycle	
2	CAR1	Share of employees commuting alone in car	
3	CAR2	Share of employees commuting two person carpool	
4	CAR3	Share of employees commuting three person carpool	
5	CAR4	Share of employees commuting four person carpool	
6	CAR5	Share of employees commuting five person carpool	
7	CAR6	Share of employees commuting six person carpool	
8	VAN_CUTR	Share of employees commuting in van	
9	BUS	Share of employees commuting in bus	
10	TRANSIT	Share of employees commuting using transit	
11	WALK	Share of employees commuting walking	
12	BIKE	Share of employees commuting using bike	
13	TELECOMMUTE	Share of employees telecommuting	
14	CWW336	Share of 3/36 CWW days off	
15	CWW440	Share of 4/40 CWW days off	
16	CWW980	Share of 9/80 CWW days off	
17	TARGET_VTR	Target VTR required (3 zones)	
18	BFL	Passenger Loading Areas	FACILITY_AMENITIES
19	BFO	Other Facility Improvements	FACILITY_AMENITIES
20	BFP	Preferential Parking Areas	FACILITY_AMENITIES
21	BFR	Bike Racks and Bike Lockers	FACILITY_AMENITIES
22	BFS	Shower and Lockers	FACILITY_AMENITIES
23	BGA	TMA/TMO Provided Guaranteed Return Trip	RIDE_HOME
24	BGC	Company Vehicle Guaranteed Return Trip	RIDE_HOME
25	BGE	Emergencies Guaranteed Return Trip	RIDE_HOME
26	BGO	Other Guaranteed Return Trip Program	RIDE_HOME
27	BGR	Rental Car Guaranteed Return Trip	RIDE_HOME
28	BGT	Taxi Guaranteed Return Trip	RIDE_HOME
29	BGU	Unscheduled Overtime Guaranteed Return	RIDE_HOME
30	BHF	Flextime for Ride sharers (Work Shifts)	FLEXTIME
31	BHG	Flextime for Ride sharers (Grace Period)	FLEXTIME
32	BMC	Commuter Information Center	MARKETING
33	BMF	Commuter Fairs (Marketing)	MARKETING
34	BMG	Focus Groups (Marketing)	MARKETING
35	BMM	Posted Materials (Marketing)	MARKETING
36	BMN	New Hire Orientation (Marketing)	MARKETING
37	BMO	Other Marketing Elements	MARKETING
38	BMP	Personal Communication (Marketing)	MARKETING
39	BMR	Company Recognition (Marketing)	MARKETING

No.	Field Name	Description	Grouping
40	BMS	Special Interest Club (Biking, Walking)	MARKETING
41	BMT	TMA/TMO Membership (Marketing)	MARKETING
42	BMW	Written Materials (Marketing)	MARKETING
43	BMZ	ZipCode Meetings (Marketing)	MARKETING
44	BRC	Regional Commuter Management Agency	RS_MATCH
45	BRE	Employer-Based Rideshare Matching System	RS_MATCH
46	DA	Transportation Allowances	FINANCIAL
47	DFB	On-Going Bike-to-Work Subsidies	FINANCIAL
48	DFC	On-Going Carpooling Subsidies	FINANCIAL
49	DFI	Introductory Transit Passes or Subsidies	COMMTAX
50	DFO	Other Direct Financial Subsidies	FINANCIAL
51	DFS	Subsidized Vanpool Seats	COMMTAX
52	DFT	On-Going Transit Subsidies	COMMTAX
53	DFV	On-Going Vanpooling Subsidies	COMMTAX
54	DFW	On-Going Walk-to-Work Subsidies	FINANCIAL
55	DNA	Auto Services (Fuel, Oil, Tune-Up)	DIRECT_NONFINANCIAL
56	DNC	Gift Certificates	DIRECT_NONFINANCIAL
57	DNF	Free Meals	DIRECT_NONFINANCIAL
58	DNO	Other Direct Non-Financial Incentives	DIRECT_NONFINANCIAL
59	DNP	Catalogue Points	DIRECT_NONFINANCIAL
60	DNT	Additional Time Off with Pay	DIRECT_NONFINANCIAL
61	DPC	Increased Parking Costs for Drive Alones	PARKMGT$
62	DPO	Other Parking Management Strategies	PARKMGT$
63	DPS	Subsidized Parking for Ridesharers	PARKMGT$
64	DTH	Work at Home (TeleCommuting)	TELE
65	DTS	Work at Satellite Center (TeleCommuting)	TELE
66	DW_3	3/36 Compressed Work Week Schedule	COMPRESSED
67	DW_4	4/40 Compressed Work Week Schedule	COMPRESSED
68	DW_9	9/80 Compressed Work Week Schedule	COMPRESSED
69	DWO	Other Compressed Work Week Schedule	COMPRESSED
70	IBO	Other Employee Benefits and Services	ONSITE
71	IBP	Drawings, Free Meals, Certificates, etc.	DIRECT_NONFINANCIAL
72	IBV	Company Owned/Leased Vanpools	VANPOOL
73	ISC	On-Site Childcare Service	ONSITE
74	ISO	Other On-Site Services	ONSITE
75	ISS	Cafeteria, ATM's, Postal, Fitness Center	ONSITE
76	IST	Transit Information or Pass Sales	ONSITE
77	OOO	Other Not Classified by Other Codes	UNCLASSIFIED
78	XXX	Incentives not Required	NOINCENTIVES
79	FACILITY_AMENITIES	Grouped incentives	
80	RIDE_HOME	Grouped incentives	
81	FLEXTIME	Grouped incentives	
82	MARKETING	Grouped incentives	
83	RS_MATCH	Grouped incentives	
84	FINANCIAL	Grouped incentives	
85	PARKMGT$	Grouped incentives	
86	TELE	Grouped incentives	
87	COMPRESSED	Grouped incentives	
88	VANPOOL	Grouped incentives	
89	ONSITE	Grouped incentives	
90	DIRECT_NONFINANCIAL	Grouped incentives	
91	OTHER	Grouped incentives	
92	COMMTAX	Grouped incentives	
93	VTR_CUTR	Vehicle trip rate	
94	EMP_TRIPS_CUTR	Total Employee trips	
95	VEH_TRIPS_CUTR	Total Vehicle trips	
96	Delta_VTR_CUTR	Change in vehicle trip rate between two years	

TUCSON DATA

The Tucson data was obtained from Pima County Association of Governments and consisted of 1,438 total records from 317 company worksites. Each Tucson record was similar to a Los Angeles data record in that it contained information related to worksite characteristics such as number of employees, shares of the different modes of transportation used by employees for commuting; the different incentives offered etc. Each record included a total of 58 fields, which are shown in Table 2. Using the information about the incentive groupings for Los Angeles data, several Tucson incentives were combined together in similar compatible Los Angeles grouped incentives. The data collected was from year 1996 to 2001. Some worksites collected data only for a subset of these years. The vehicle trip rate (VTR) for each worksite was calculated from the given data using the formula -

VTR = 100 * (DriveAlone + (CarPool + VanPool)/2.19) / (DriveAlone + Bus + CarPool + VanPool + Walk + Bike + Motorcycle + CWW336 + CWW440 + CWW980)

Where,
DriveAlone – Number of employees driving alone
Motorcycle – Number of employees commuting by motorcycle
CarPool – Number of employees commuting using carpool
VanPool – Number of employees commuting in van
BUS – Number of employees commuting by bus
TRANSIT – Number of employees commuting using transit
WALK – Number of employees commuting walking
Bike – Number of employees commuting by bike
CWW336 – Number of 3/36 days off
CWW440 – Number of 4/40 days off
CWW980 – Number of 9/80 days off
2.19 – average carpool & vanpool occupancy

The change in VTR (Delta_VTR) was calculated similar to the way it was calculated for Los Angeles data and the last year records were removed.

Table 2: Tucson Data variable description and grouping

No.	Field	Description	Grouping
1	PermId	Id number of the employer	
2	PlanYr	Plan year	
3	AloneShare	Alone share	
4	BusShare	Bus share	
5	CVpoolShare	Carpool + Vanpool share	
6	WalkShare	walk share	
7	McycleShare	Motorcycle share	
8	BcycleShare	Bicycle share	
9	HiMiles	Average Miles traveled	
10	HiMinute	Average minutes traveled	
11	CWW336	3/36 compressed work week share	
12	CWW440	4/40 compressed work week share	
13	CWW980	8/80 compressed work week share	
14	VTR	Vehicle trip rate (in fraction)	
15	NoEmp	Number of employees	
16	AC	Alternate Mode Information	MRKT
17	AQ	Post Air Quality Information	MRKT
18	AW	Adjusted Work Hours	FLEX
19	BP	Bus Pass Sales on Site	ONSITECONV
20	BR	Bicycle Racks	FACILITY_AMENITIES
21	BS	Bus Subsidy	COMMTAX
22	BV	Busing Vehicle	VANPOOL
23	CC	Matching Service	RSMATCH
24	CD	Covered Parking	FACILITY_AMENITIES
25	CG	Alternative Fuel Vehicles	VANPOOL
26	CP	Carpool Subsidy	FINANCIAL
27	CS	Coordination with Transit Provider	MRKT
28	CV	Carpooling Vehicle	VANPOOL
29	CW	Compressed Work Week	CWW
30	DC	Daycare Facilities on Site	ONSITECONV
31	DQ	Dissemination of Air Quality Information	MRKT
32	DW	Drawing for Prizes	MRKT
33	ES	Employee Shifts Between Sites	TELE
34	EV	Alternate Mode Campaign	MRKT
35	FP	Fee for Parking	PARKMGT$
36	FW	Field Worker	UNCLASSIFIED
37	GP	Guaranteed Ride Home Program	GRH
38	IC	Information Center	MRKT
39	IN	Incentives for Employees to Live Close	FINANCIAL
40	IP	Incentive Programs	DIRECT_NONFIN
41	MP	Map Board	RSMATCH
42	NE	New Employee Information	MRKT
43	NL	Newsletter Articles	MRKT
44	PP	Preferred Parking	FACILITY_AMENITIES
45	RC	Rideshare Committee	MRKT
46	RP	Rebate not to Use Parking	PARKMGT$
47	SA	Staging Area	FACILITY_AMENITIES
48	SC	Carpool Campaign	MRKT
49	SK	Speakers	MRKT
50	SV	Shuttle Vehicle	VANPOOL
51	SW	Showers/Lockers	FACILITY_AMENITIES
52	TF	Transportation Fair	MRKT
53	VC	Vanpooling Vehicle	MRKT
54	VP	Vanpool Subsidy	COMMTAX
55	VV	Vanpooling Vehicle	VANPOOL
56	WB	Walking Campaign	MRKT
57	WH	Work at Home	TELE
58	WS	Bicycle Campaign	MRKT
59	Delta_VTR	Change in VTR	Change in VTR

WASHINGTON DATA

The Washington data was obtained from Washington State Department of Transportation and consisted of 2,482 total records from 1,038 company worksites. Each record represents information from a specific company worksite for a specific year. The information in a record includes worksite characteristics such as number of employees, the type of the company, shares of the different modes of transportation used by employees for commuting, the different incentives, and the preferences of the employees for the incentives they feel important. Each record included a total of 106 fields, which are shown in Table 3. The table also shows the grouping of the similar incentives as done for other datasets. For Washington, bi-yearly data were collected from 1995 to 2001. The vehicle trip rate for each worksite was calculated from the given data using the formula

VTR = 100 * (Q4_ALONE + Q4_CAR2/2 + Q4_CAR3/3 + Q4_CAR4/4 + Q4_CAR5/5 + Q4_CAR6/6 + Q4_VAN7/7) / (Q4_ALONE + Q4_BIKE + Q4_BUS + Q4_CAR2 + Q4_CAR3 + Q4_CAR4 + Q4_CAR5 + Q4_CAR6 + Q4_OTHER + Q4_TELE + Q4_VAN7 + Q4_WALK + Q5_DAYS3*2 + Q5_DAYS4 + Q5_DAYS7*1.5+ Q5_DAYS9*0.5))

Where,
Q4_ALONE – Number of employees driving alone
Q4_CAR2– Number of employees commuting two together
Q4_CAR3 – Number of employees commuting three together
Q4_CAR4 – Number of employees commuting four together
Q4_CAR5 – Number of employees commuting five together
Q4_CAR6 – Number of employees commuting six together
Q4_VAN7 – Number of employees commuting in van
Q4_BUS – Number of employees commuting by bus
Q4_WALK – Number of employees commuting walking
Q4_BIKE – Number of employees commuting by bike
Q4_TELE – Number of employees telecommuting
Q5_DAYS3 – Number of employees on 3/36
Q5_DAYS4 – Number of employees on 4/40
Q5_DAYS7 – Number of employees on 9/80
Q5_DAYS9 – Number of employees on 9/80

The change in VTR (Delta_VTR) was calculated similar to the way it was calculated for Los Angeles data and the last year records were removed.

Table 3: Washington Data variable description and grouping

No.	Field Name	Description	Grouping
1	CTRID	CTRID	
2	AnnReYR	Plan Year	
3	NonProf	Non-profit organization	
4	Agri	Agriculture organization	
5	Finance	Finance organization	
6	InfoServ	Info services organization	
7	Health	Health organization	
8	Retail	Retail organization	
9	Manufact	Manufacturing organization	
10	Services	Services organization	
11	pubUtil	Public utilities organization	
12	Construc	Construction organization	
13	Xport	Transportation organization	
14	Govern	Government organization	
15	Other	Other organization	
16	Offer2Al	Offered to All	
17	Union	Union	
18	Shifts	Shifts	
19	OnParkSp	Onsite Parking Spaces	
20	OfParkSp	Offsite Parking Spaces	
21	LOnParkP	Leased Onsite Parking Price	
22	LOfParkP	Leased Offsite Parking Price	
23	OwnOnPrk	Own Onsite Parking	
24	OnPrkChr	Onsite Parking Charge	
25	OwnOfPrk	Own Offsite Parking	
26	OfPrkChr	Offsite Parking Charge	
27	PyPrkChr	Pay Parking Charge	
28	OnOfPsub	On off parking sub	ONSITE
29	FrePrkQM	Free Parking 1/4 mile	
30	ETCAdTrn	ETC Additional Training	ONSITE
31	ETCOnsit	ETC Onsite	ONSITE
32	DstrInfo	Distribute Info	
33	PstMatrl	Post Materials	
34	Orientat	CTR Orientation	
35	CTREvent	CTR Events	MRKT
36	CTREmail	CTR E-mail	MRKT
37	Articles	Articles	MRKT
38	RideMApp	Ride match Apps	RS_MATCH
39	WPycheck	With Paychecks	MRKT
40	Drawings	Drawings	DIRECT_NONFINANCIAL
41	Leave	Leave	DIRECT_NONFINANCIAL
42	OthPromo	Other Promo	MRKT
43	CovBikeN	Covered Bike Number	FACILITIES_AMENITIES
44	UncovBkN	Uncovered Bike Number	FACILITIES_AMENITIES
45	LockersN	Lockers Number	FACILITIES_AMENITIES
46	ShowersN	Showers Number	FACILITIES_AMENITIES
47	ShelterN	Shelters Number	FACILITIES_AMENITIES
48	OthrAmeN	Other Amenities 1 Number	FACILITIES_AMENITIES
49	CP-SpacN	Carpool Spaces Number	FACILITIES_AMENITIES
50	VP-SpacN	Vanpool Spaces Number	FACILITIES_AMENITIES
51	SOVPkch	SOV Parking Charge	PARKMGT
52	SOVPkChN	SOV Parking Charge Number	PARKMGT
53	RedSOVPN	Reduced SOVP Number	PARKMGT
54	TransSub	Transit Subsidy	COMMTAX
55	FerrySub	Ferry Subsidy	FINANCIAL
56	VanPSub	Vanpool Subsidy	COMMTAX
57	CarPSub	Carpool Subsidy	FINANCIAL
58	WalkSub	Walking Subsidy	FINANCIAL
59	BikeSub	Bike Subsidy	FINANCIAL
60	EmpOnFlx	Employees on flextime	FLEXTIME
61	EmpWiGRH	Employees with GRH	GRH
62	EmpInHM	Employees in-house match	RS_MATCH

No.	Field Name	Description	Grouping
63	EmpPubM	Employees public match	RS_MATCH
64	FVwrkEmp	FV work employees	
65	EmpNo	Number of Employees	
66	CWW3	Percentages of employees on 3/36 CWW	CWW
67	CWW4	Percentages of employees on 4/40 CWW	CWW
68	CWW5	Percentages of employees on 5/40	
69	CWW7	Percentages of employees on 7/40 CWW	CWW
70	CWW9	Percentages of employees on 9/80 CWW	CWW
71	CWWOTHER	Percentages of employees on other CWW	CWW
72	AloneShr	Alone share	
73	BikeShr	Bike share	
74	BusShr	Bus share	
75	CarShr	Cars hare	
76	OtherShr	Other share	
77	TeleShr	Tele share	
78	VanShr	Van share	
79	WalkShr	Walk share	
80	DSvTC2wk	Days saved telecommuting in two weeks	
81	AdminSh	Administration job Share	
82	CrPrLaSh	Craft/Production/Labor Share	
83	MgntSh	Management job Share	
84	SaleMktSh	Sales/Marketing job Share	
85	CstSrvSh	Customer Service job Share	
86	OtherSh	Other job Share	
87	ProTecSh	Professional/Technical job Share	
88	PEmpPCar	Prefer provide car for work	
89	PTrlunch	Employee Prefer Transport during lunch	
90	PreGRH	Employee Prefer GRH	
91	PflxCVpB	Employee Prefer flex to meet CVpool bus	
92	PFinance	Employee Prefer financial incentive	
93	PDsCVpSp	Employee Prefer reserved discounted CVpool space	
94	PPrlHCVp	Employee Prefer Personalized help for CVpool	
95	PCovbkPk	Employee Prefer covered bicycle parking	
96	PLckAShw	Employee Prefer lockers & showers	
97	POnChild	Employee Prefer onsite childcare	
98	PreCWW	Employee Prefer CWW	
99	PreTele	Employee Prefer to telecommute	
100	Pimptran	Employee Prefer improved access to transit	
101	SOV	SOV	
102	VMT	VMT	
103	VTR	VTR	
104	CBD	Central Business District	
105	suburb	Suburban area	
106	outside	Outside suburban area	
107	DeltaVTR	DeltaVTR	

III. MODEL BUILDING APPROACHES

The above cleaned data and its subsets were used in the model building process. Two different approaches were used in model building: non-linear neural networks and linear statistical regression models. Regression models identify the simple linear relationships between the independent inputs and the dependent output variable, whereas neural network models can explain much more complex non-linear relationships between the independent inputs and dependent variables. Therefore, the results provided by the regression models can be used as the baseline for the comparison of the various neural network models.

NEURAL NETWORKS

Neural networks are essentially a group of highly interconnected and relatively simple computational units, as illustrated in Figure 1. Each of these computational units performs processing of its inputs to produce a single output. The neural network connects the output of each unit to the inputs of many other units through different weights.

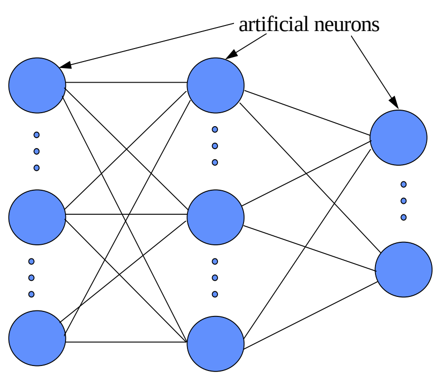


Figure 1: Neural network consisting on interconnected neuron

Figure 2 shows a typical artificial neuron (i.e., any single individual neuron) which adds all of its weighted inputs and uses a sigmoid output function to generate its output. Other output functions, such as the hyperbolic tangent, are sometimes used. As can be seen in Figure 2, neural network learning consists of finding the correct weighting of the input data so that the model of best fit may be found.

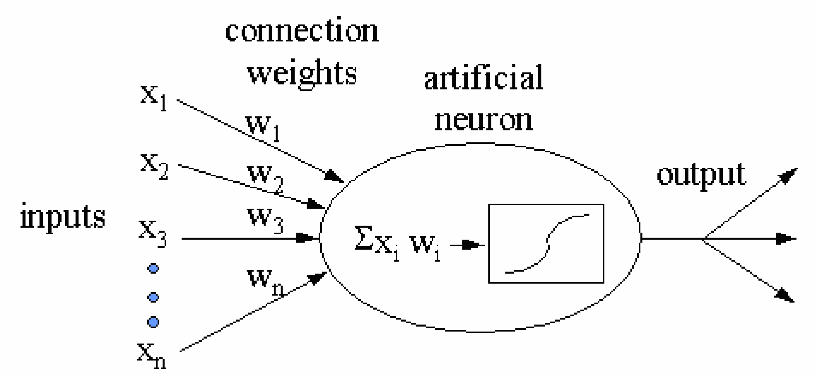


Figure 2: Artificial neuron with inputs and outputs

To help understand the process followed to build the model, defining several key terms is necessary.

Training is a process that uses one of several learning techniques to modify the weights in an orderly fashion. The *training set* of data is a list of paired input and desired output patterns used in supervised training. All of the information the network needs to learn must be in the training set. The inputs can be numbers, symbols, or pictures.

The *testing set* is an extract of the training set used while building the model to prevent over-fitting. Over-fitting the training data can occur when the neural network produces a nonlinear model that fits the training data perfectly, but fits the test data very poorly. The goal is to fit both the training and testing data with approximately the same overall error. Therefore, the testing data set is used to analyze the model's ability to interpolate the training/testing data regularly during training. Training is halted when the test performance starts to degrade.

The *validation set* is independent of the training/testing set and typifies the data that will be seen by the model in the outside world.

Figure 3 illustrates the type of network that is most popular today - a multi-layer, fully-connected, feed forward neural network. These types of networks consist of two or more layers of individual neurons. Each neuron in a given layer receives inputs from all the neurons in the previous layer. The output of the previous layer is input to all neurons in the succeeding layer. The middle layers, between the input and output layers are called hidden layers since they are not directly accessible. The sigmoid or the hyperbolic tangent allows each computational unit to implement a nonlinear mapping between its inputs and output. This allows the networks to model nonlinear relationships that may exist in the data. Once this network is trained, it will produce the desired m-dimensional output given an n-dimensional input. The records provide the desired input-output associations used to train the network.

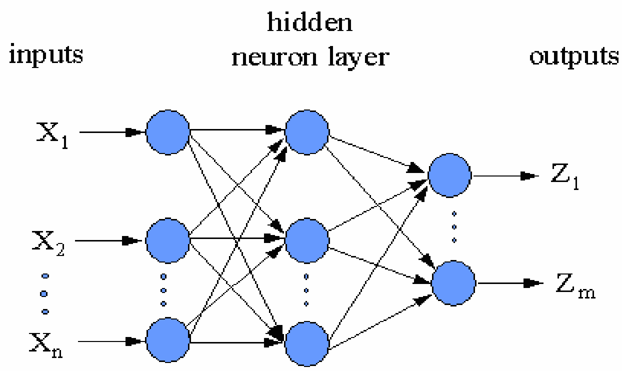


Figure 3: Typical multi-layered fully connected feed forward neural network

The popularity of these types of networks is mostly due to two factors. First, a two-layer feed forward network is capable of implementing any association between inputs and outputs. The second factor that accounts for the popularity of these types of networks is the existence of a well-defined training method called back propagation. This training method can find the weight values that will allow the neural network to produce the desired m-dimensional output given an n-dimensional input.

Before training a network using the back propagation method, the network builder must identify inputs and outputs. This is a critical step for building an accurate neural network and should be done by someone aware of the application domain. During training, that portion of a record identified as input is presented to the network. If the output of the network differs from the output portion of the record, then the neural network changes the weights of the network. The back propagation method specifies what changes to make to the weights so that the neural network reduces the difference between the actual and the desired network output. All of the training records are presented to the network, and corrections made. This makes up one training cycle. Typically, training a network requires many training cycles until the cumulative errors of all records in one cycle are below an acceptable level. Thus back propagation uses the records in the training data repeatedly to change the weights in the network so that the difference between the desired and the actual network output is below an acceptable level.

General Network Structure

For the reasons mentioned above, the type of neural network selected to predict the change in VTR was a multi-layer, fully-connected, feed forward type. Neural network model builders have applied these types of networks successfully for prediction and classification problems in a variety of fields.

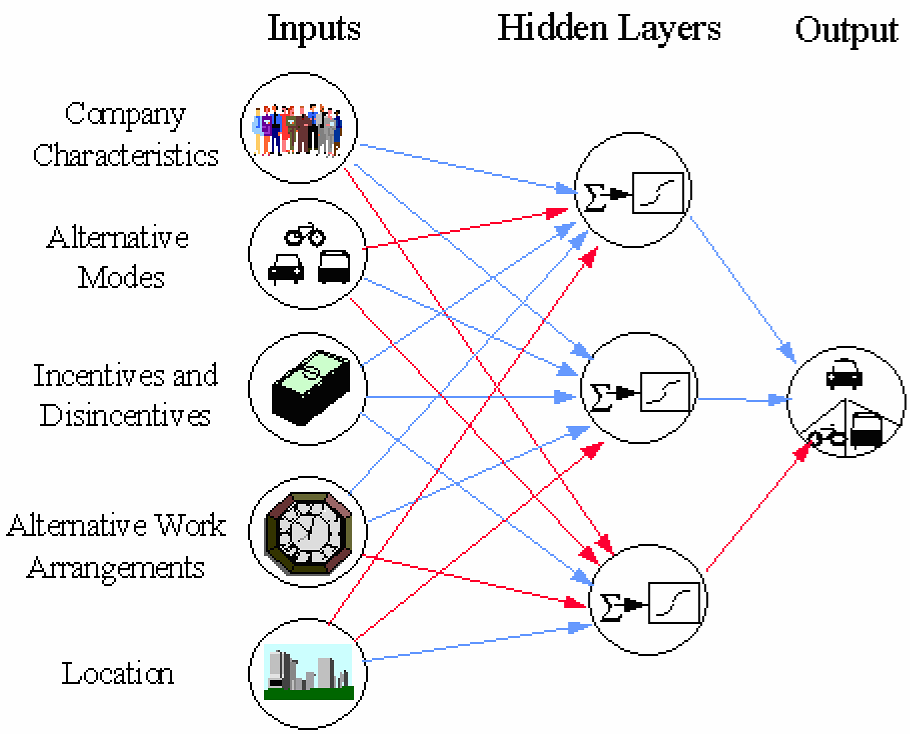


Figure 4: Network structure for Worksite trip reduction model

Figure 4 shows an overview of the network's structure. All of the inputs to the network and the output have been described in this report. The goal of this network is to predict the change in VTR that a company will obtain due to the specific combination of incentive plans selected by the company and the specific characteristics of that company. The neural network determines the number of hidden nodes and layers during training and will be different from those illustrated.

Software Used to Build the Neural Networks

Neuralware's NeuralWorks Predict v5.0, a plug-in in Microsoft's Excel was used for building the neural network models. Microsoft's database program, Access, was also used to manipulate the data before training the network. The hardware used was a 1 GHz Pentium III PC with 512 Mbytes of RAM.

Predict simplifies the different aspects of the neural network training process by allowing the network builder to select many parameters that can affect the performance of the final model. Researchers were able to customize the following key parameters:

1. The types and maximum number of computational units to use.
2. The type of error evaluation function. For example, one error evaluation function may emphasize higher correlation between predicted and desired network output, while another may emphasize greater accuracy of the predicted quantity.
3. Whether to consider noise (i.e., errors in the data caused by individual differences between companies and people) in the training data.
4. In what proportion to break up the data used for training the network and the data used to test the trained network.
5. Whether to eliminate as input to the network those input variables that had little correlation with the output variable.

REGRESSION ANALYSIS

Regression models were built using the SPSS system, release 11.1 for Windows.

To provide a baseline of comparison of the ability of the neural network model to predict changes in VTR correctly, an independent model was first created by means of factor analysis and stepwise regression.

Initial regressions suffered from multi-collinearity within the data. Since many independent variables were inter-correlated, a possibility exists that the coefficients resulting from model runs would not fully reflect the effects of each of the independent variables. The initial approach to eliminating the effects of the multi-collinearity was to run a factor analysis.

Generally, factor analysis is used as a data reduction technique. The analytical procedure involves creating uncorrelated (orthogonal) combinations of the initial dependent variables.

In common practice, the purpose of the analysis is to reduce a mass of variables to a reasonable number of elements (e.g., 10), which the analyst can understand and explain. Often, the selection of factors to use is limited to those that explain at least as much variance as an independent variable, i.e., the output factor has an eigenvalue of at least one.

However, here, the purpose was to create a series of uncorrelated factors to use as independent variables in regression analysis, without attempting to explain the meaning of the factors. When factor analysis is used in hard science applications, factors are commonly retained to the point where 95 percent of the variance in the data is explained. Some factor analysts even maintain that any factor with a positive eigenvalue is relevant for analysis (Hair, 1984). Since a stepwise regression procedure was to be used later for model building, CUTR researchers decided that the stepwise procedure would determine the significance of the factors produced by the analysis, rather than limiting the factors output through analysis of eigenvalues.

The stepwise regression was set to accept variables that significantly improved the model at a 95 percent confidence level. When the analysis had been completed, the factors were then reconverted into the original component independent variables. The conversion was made by multiplying the coefficients assigned by the regression model to the factors by the matrix of the factor loadings of the original variables. The resulting equations predicted the change in VTR.

No reduction in the number of predictor variables was obtained by using this approach. In fact, compared to stepwise regressions, the number of predictor variables was larger because nearly every input variable loaded onto a factor at some point. Even reducing the factors used to those with eigenvalues above 1.0 would not have helped this situation.

The second step taken to resolve this difficulty was to remove correlated variables. While this step risked reducing the predictive power of the model, the resulting coefficients would be much more easily interpretable. Also, a stepwise selection of variables reduces the data required to make a prediction. This should eventually reduce the burden placed on model users.

The actual finding was that taking this step only marginally diminished the explanatory power of the model. The overall predictive accuracy of the results, as shown in later sections, was only slightly reduced no matter the method of regression model building.

In order to understand the results listed in later chapters, one must understand the methods used in SPSS regression. The stepwise methodology has been briefly explained previously, but the forward, backward, forced enter, and stepwise will be discussed further so that the reader will understand the processes involved in the varying methodologies of model building via regression.

Forward regression is very much like stepwise regression in that SPSS will first compute which predictor variable has the largest bivariate correlation with the dependent variable. From this point, the SPSS program will then add in various other variables which explain that largest amount of variance and these additional variables will only be included if they explain a significant amount of additional variance. The forward regression process stops when there are no more variables remaining that can explain a significant amount of the variance.

In the case of backward regression, the program enters all of the independent variables and then removes them one by one, based upon a preset significance value. The default value in this case removes all variables which don't explain at least 90 percent of the variance. The backward regression process stops when there are no more variables remaining for removal that meet this requirement.

Stepwise regression combines both forward and backward methodologies. It is a complex procedure in that it will remove or add variables as the inter-correlations change between variables already in the regression formula. That is, a variable which significantly explained the variance in the dependent variable is weakened by the addition of another variable and is therefore removed by the program. Stepwise is the most popular of all regression methods.

Forced enter regression method is more simple than stepwise, forward, or backward, in that the program simply accepts all of the specified variables and builds a regression equation regardless of the significance levels of the entered independent variables.

MEASURING THE PERFORMANCE OF TRAINED MODELS

As stated earlier, the dataset selected for building a neural network/regression models is divided in two disjoint sets "training/testing set" and "validation set". In case of neural network models, the training/testing set is once more split into two separate sets, the "training set" and the "testing set". The training set is used to train the network while the testing set is used to test the performance of the network as it is being trained and control the number of hidden units in the network. By default, Predict uses 70 percent of the random data from training/testing set as the training set and remaining 30 percent as the test set, although the network builder can change these values to any other proportions. The whole training/testing set is used for training by the statistical software for building the regression models. The validation set which is a representative sample of the original input data is used in the final performance evaluation of the models. The neural network/statistical software does not use the validation set in building the model.

To get some raw estimates of the accuracies of the predicted Delta_VTR (change in VTR), the range of the continuous variable Delta_VTR is discretized into 8 bin/classes so that each bin contains approximately $1/8^{th}$ of the data. Since most of the changes in VTR are small from one period to the next, the bin classification approach to assessing the model performance helps focus on plans with large changes in VTR. Also since each bin has to contain approximately same amount of data, the bin ranges used for the three datasets are not consistent. Two accuracies, the "Exact Accuracy" and the "One-off Accuracy" measures are used.

The “Exact Accuracy” measures, for how many records the predicted bin was exactly same as actual bin. The “One-off Accuracy” measures, for how many records the predicted bin was same as actual bin or one of its neighboring bins.

In the next section, we do the comparison of the various neural network and regression models using these three performance measures

The first performance measure is the ‘Bin Classification Accuracy on moderate range of change in VTR (i.e. bins a2 to a5)’. The second performance measure is the ‘Bin Classification Accuracy on full range of change in VTR (i.e. all bins)’ and the third measure is the R-square between the actual Delta_VTR and the Predicted Delta_VTR. R-square is a statistical term for the amount of variance for which the regression formula is able to account.

The formula used for calculating R-square is

$$r = \frac{\sum (x-\overline{x})(y-\overline{y})}{\sqrt{\sum (x-\overline{x})^2 \sum (y-\overline{y})^2}}$$

Where,
x = Actual Delta_VTR
y = Predicted Delta_VTR

It has to be kept in mind that the first two performance measures, which assess the accuracy over a distribution, may be more useful than the R-square variance measure. It was found that often the models had a poor R-square but better Bin Classification Accuracies on moderate and full range of change in VTR.

DATASETS MANIPULATIONS

Various data manipulations were performed on the different datasets to improve the performance of the models built on them. The individual incentives in the datasets were replaced by the grouped incentives, thus reducing the total number of variables in the dataset. This helped in reducing the complexity and adding more explanatory power to of the models. In some cases the data in bins lying in moderate range of change in VTR was over-sampled, meaning that the samples in those bins were duplicated to give more importance to these samples as against to the data in other ranges of change in VTR was under-sampled, meaning that some samples in these bins were dropped to reduce their importance. Models were also built on the data which was manipulated by both grouping of incentives and over/under sampling of data in some bins.

IV. MODEL BUILDING

LOS ANGELES MODELS

Phase I: Los Angeles Full Sample Data

The Los Angeles dataset consisted of 25,459 total examples. For datasets from other areas, ten percent of the dataset was left aside as the validation set while the rest of the data became the training/testing set. However, because of the large size of the Los Angeles dataset, twenty percent of the data was left aside as the validation set. The ranges of the eight bins based on the change in VTR values and the number of examples in each bin for validation and training/testing set are shown in Table 4.

Table 4: Los Angeles Full Sample Data – No. of Records in bins

range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)							
		> - 7	[- 7 to - 4)	[- 4 to - 2)	[- 2 to -0)	[-0 to 1)	[1 to 2.5)	[2.5 to 5)	5 >=
Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**
Validation	**2537**	347	286	356	423	237	264	262	362
Training	**22922**	3087	2969	2943	3988	2155	2260	2612	2908

A first attempt to build neural network models using all the variables present in the data resulted in very poor performance. So, different linear regression modeling approaches were used to select different sets of variables. Forced Enter, Stepwise, Forward and Backward regression approaches yielded different subsets of variables as indicated in the table below. Forced Enter consists of entering all variables to build the model, regardless of significance in explaining variance. Forward method involves the entry of variables one at a time, based on the significance of each variable. Backward method involves entering all variables and removing each one by one, according to a preset significance value. Finally, Stepwise method utilizes both forward and backward methods, obtaining the best set of variables for the regression equations.

After the different methodologies were applied to building the SPSS regression models, the neural network models were built using each of the different set of variables obtained from each of the four regression building methods, forced enter, stepwise, forward, and backward. Table 5 shows the different variables selected by each of the different SPSS regression methods. These variables were then used to train the neural net models.

Table 5: Variables selected by different regression methods

Names	Variables selected using regression method			
	Force Enter	stepwise	Forward	Backward
Share employees commuting by motorcycle	√	√	√	
Share employees commuting alone in car	√	√	√	
Share employees commuting two together	√	√	√	√
Share employees commuting three together	√	√	√	√
Share employees commuting four together	√			√
Share employees commuting five together	√	√	√	
Share employees commuting six together	√			
Share employees commuting in van	√			√
Share employees commuting in bus	√			√
Share employees commuting using transit	√			√
Share employees commuting walking	√			√
Share employees commuting using bike	√	√	√	√
Share employees tele-commuting	√	√	√	√
Share of 3/36 CWW days off	√	√	√	√
Share of 4/40 CWW days off	√	√	√	√
Share of 9/80 CWW days off	√	√	√	√
Target VTR required (3 zones)	√	√	√	√
Passenger Loading Areas	√			
Other Facility Improvements	√	√	√	√
Preferential Parking Areas	√			√
Bike Racks and Bike Lockers	√			√
Shower and Lockers	√			
TMA/TMO Provided Guaranteed Return Trip	√			
Company Vehicle Guaranteed Return Trip	√	√	√	√
Emergencies Guaranteed Return Trip	√	√	√	√
Other Guaranteed Return Trip Program	√	√	√	√
Rental Car Guaranteed Return Trip	√			
Taxi Guaranteed Return Trip	√	√	√	√
Unscheduled Overtime Guaranteed Return	√			√
Flextime for Ridesharers (Work Shifts)	√			
Flextime for Ridesharers (Grace Period)	√			
Commuter Information Center	√			
Commuter Fairs (Marketing)	√			
Focus Groups (Marketing)	√			
Posted Materials (Marketing)	√	√	√	√
New Hire Orientation (Marketing)	√	√	√	√
Other Marketing Elements	√	√	√	√
Personal Communication (Marketing)	√			
Company Recognition (Marketing)	√			√
Special Interest Club (Biking, Walking)	√	√	√	√
TMA/TMO Membership (Marketing)	√			
Written Materials (Marketing)	√			√
ZipCode Meetings (Marketing)	√			
Regional Commuter Management Agency	√			√
Employer-Based Rideshare Matching System	√			√
Transportation Allowances	√			
On-Going Bike-to-Work Subsidies	√			
On-Going Carpooling Subsidies	√	√	√	√
Introductory Transit Passes or Subsidies	√			
Other Direct Financial Subsidies	√			
Subsidized Vanpool Seats	√			√
On-Going Transit Subsidies	√			
On-Going Vanpooling Subsidies	√			√
On-Going Walk-to-Work Subsidies	√			√
Auto Services (Fuel, Oil, Tune-Up)	√			
Gift Certificates	√			
Free Meals	√			√
Other Direct Non-Financial Incentives	√			
Catalogue Points	√			
Additional Time Off with Pay	√	√	√	√

Names	Variables selected using regression method			
	Force Enter	stepwise	Forward	Backward
Increased Parking Costs for Drive Alones	√	√	√	√
Other Parking Management Strategies	√	√	√	√
Subsidized Parking for Ridesharers	√			
Work at Home (TeleCommuting)	√			
Work at Satellite Center (TeleCommuting)	√			
3/36 Compressed Work Week Schedule	√			
4/40 Compressed Work Week Schedule	√		√	
9/80 Compressed Work Week Schedule	√			√
Other Compressed Work Week Schedule	√	√	√	√
Other Employee Benefits and Services	√			
Drawings, Free Meals, Certificates, etc.	√			
Company Owned/Leased Vanpools	√			
On-Site Childcare Service	√			
Other On-Site Services	√			
Cafeteria, ATM's, Postal, Fitness Center	√	√	√	√
Transit Information or Pass Sales	√			
FACILITY_AMENITIES	√			√
RIDE_HOME	√	√	√	√
FLEXTIME	√			
MARKETING	√			√
RS_MATCH	√			
FINANCIAL	√			
PARKMGT$	√		√	
TELE_GRP	√			
COMPRESSED	√	√	√	√
VANPOOL	√			√
ONSITE	√			
DIRECT_NONFINANCIAL	√			
OTHER	√			
COMMTAX	√			
Vehicle trip rate		√	√	√
Total Employee trips	√			
Total Vehicle trips	√			

Figures 5, 6 & 7 show charts of the 'bin classification accuracy on moderate range of change in VTR', 'bin classification accuracy on full range of change in VTR' and the R-square values for the regression models and the corresponding neural network models built using the variable sets selected by the regression models. It can be seen from the table 4 that the number ranges are labeled below as "a1," "a2," "a3," etcetera. For example, the "a2" labeled bin covers a range of change in VTR from -7 to -4 and the "a5" labeled bin covers a range of change in VTR from 0 to 1. Therefore, the bins "a2" through "a5" would cover the change in VTR range between -7 through +1 which can be considered as a moderate range.

In Figures 5, 6, and 7 below, for each methodology (i.e. forced enter, stepwise, forward, or backward), the models labeled as "linear model" specify the SPSS regression models built on the variables obtained from the specific methodologies and the models labeled as a "sequence on three numbers" (Input neurons - hidden neurons - output neurons) specify the architecture of the neural net model also built from those same variables.

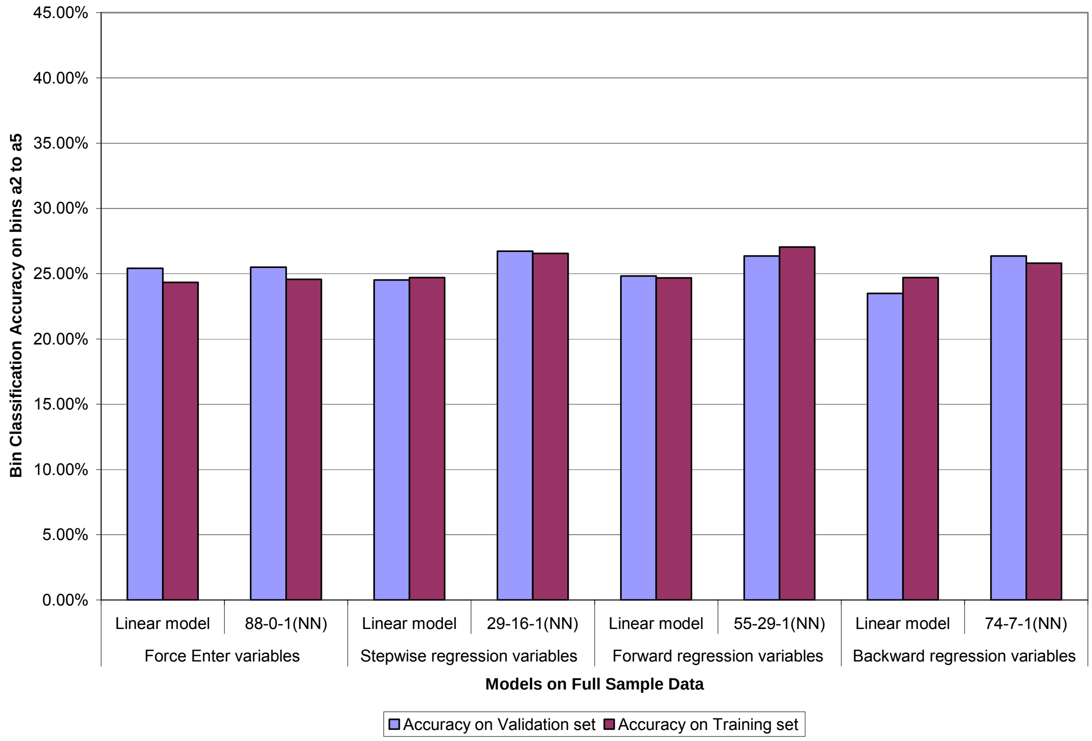


Figure 5: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on Full sample data)

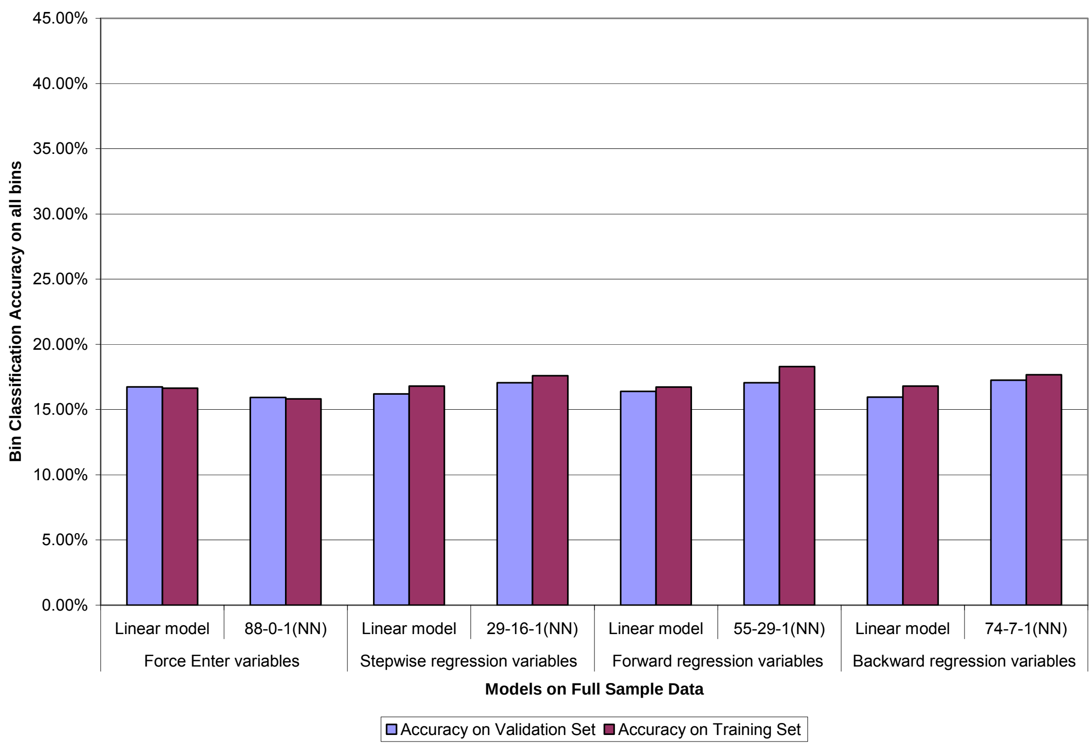


Figure 6: Bin Classification Accuracy on Full Range of change in VTR(all bins) for validation & training set (Different models on Full sample data)

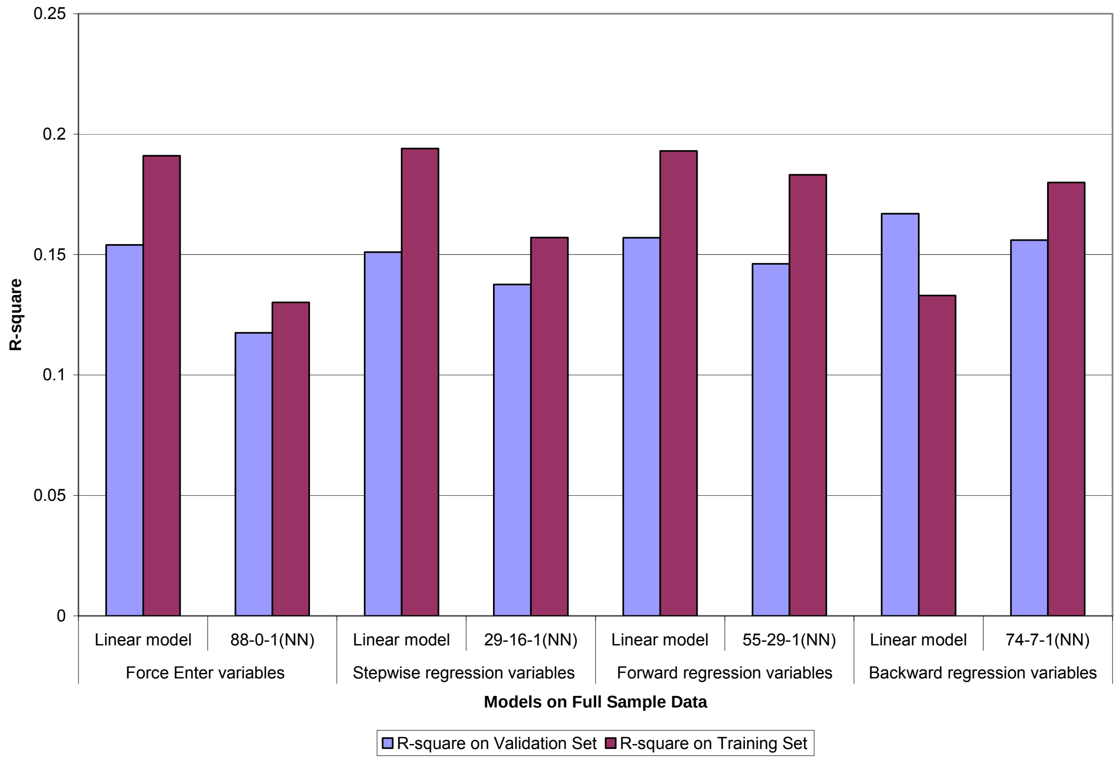


Figure 7: R-square on training and validation data (Different models on Full sample data)

It can be seen from the figures that the neural network model built using the stepwise regression variables yielded the best 'bin classification accuracy on moderate range of change in VTR', and the second best 'bin classification accuracy on full range of change in VTR'.

Table 6: Detailed accuracies for NN model on stepwise regression variables

range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)								
		> - 7	[- 7 to - 4)	[- 4 to – 2)	[- 2 to -0)	[-0 to 1)	[1 to 2.5)	[2.5 to 5)	5 >=	
Bin Number		a1	a2	a3	a4	a5	a6	a7	a8	
Validation	2537	347	286	356	423	237	264	262	362	Weighted
Training	22922	3087	2969	2943	3988	2155	2260	2612	2908	Avg on
										a2 to a5
Exact Validation	17.07%	4.61%	17.48%	27.25%	40.43%	12.66%	15.53%	9.16%	1.10%	26.73%
Exact Training	17.59%	5.54%	18.22%	25.45%	39.77%	15.03%	15.58%	9.49%	2.10%	
One-off Validation	47.22%	28.82%	45.80%	72.75%	70.45%	64.98%	35.23%	30.53%	22.93%	64.67%
One-Off Training	47.89%	29.87%	44.43%	75.03%	72.57%	68.12%	37.61%	27.41%	20.67%	

Table 6 explains how accuracy was verified on a bin by bin basis. Each of the predicted Delta_VTR obtained from the neural net and the regression models were compared against the real change in vehicle trip rate obtained from the data sets. Each of these Delta_VTR values were coded into bins using the ranges as shown in Table 4 and the results were cross tabulated to ascertain the accuracy of the predicted values versus the real values. The bin accuracy system was explained in the previous section, but for clarity, the designation "Exact Validation" refers to the accuracy on training set where the predicted bins exactly matched the actual bins. "One-off Training" refers to the accuracy on training set where the predicted bin matched the actual bin or one of the adjacent bins. "Exact Validation" refers to the accuracy on validation set where the predicted bins exactly matched the actual bins and "One-off Validation" refers to the accuracy on validation set where the predicted bin matched the actual bin or one of the adjacent bins. For example, the model shown in Table 6 accurately predicted 72.75 percent of the records showing a reduction from 0 to -7 vehicle trips.

In Figure 8 which shows the scatter plot for this model on the validation set, the upper right quadrant shows positive predicted values that matched with positive actual values and the lower left quadrant shows negative predicted values that matched with the negative actual values. The upper left and lower right represent predicted values which were mismatched to the actual values with respect to the sign (i.e., negative or positive) value.

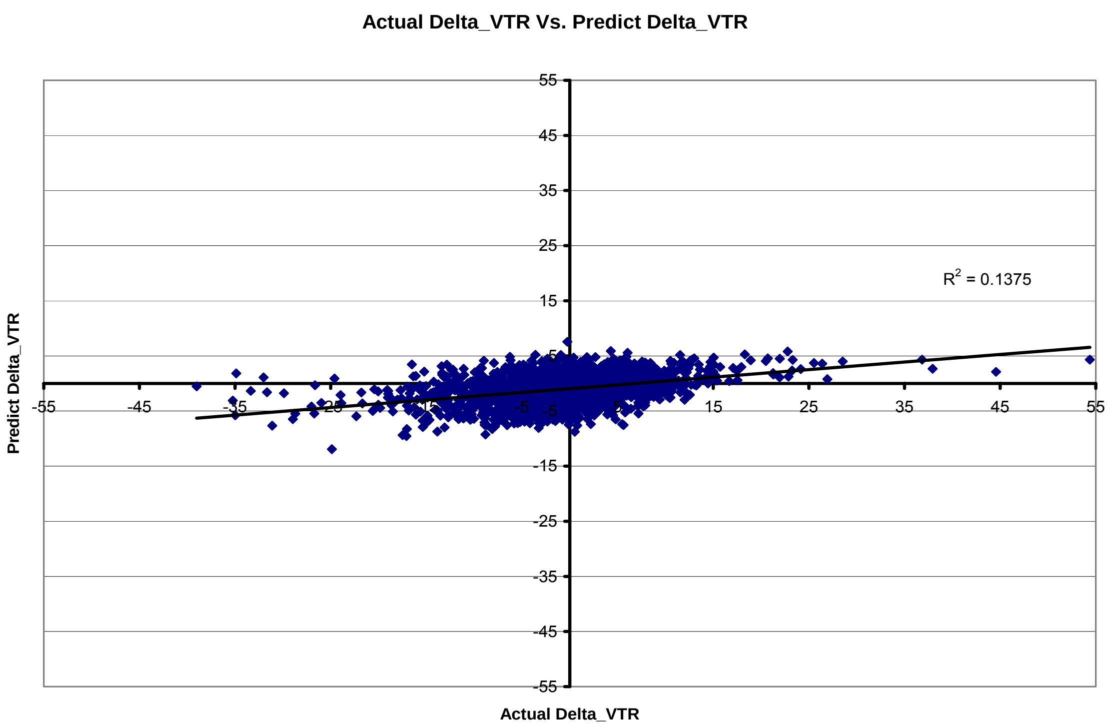


Figure 8: Scatter plot for NN model on stepwise regression variables

Phase II: Los Angeles Over-Sampled Data

To get better accuracies over the required range on change in VTR, the examples in some bins were over-sampled and others sub-sampled. Table 7 shows the new number of examples in training set due to over-sampling and sub-sampling.

Table 7: Los Angeles Over-Sampled Data – No. of Records in bins

range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)							
		> - 7	[- 7 to – 4)	[- 4 to – 2)	[- 2 to -0)	[-0 to 1)	[1 to 2.5)	[2.5 to 5)	5 >=
Bin Number		a1	a2	a3	a4	a5	a6	a7	a8
Validation	2537	347	286	356	423	237	264	262	362
Training	35227	4564	2534	6629	11500	4000	3320	1324	1356

The neural network models were built on variables selected by forward and stepwise regression. Additionally, the costs associated with each incentive were added for each of the respective stepwise incentives and a neural network model was constructed from these cost-modified variables.

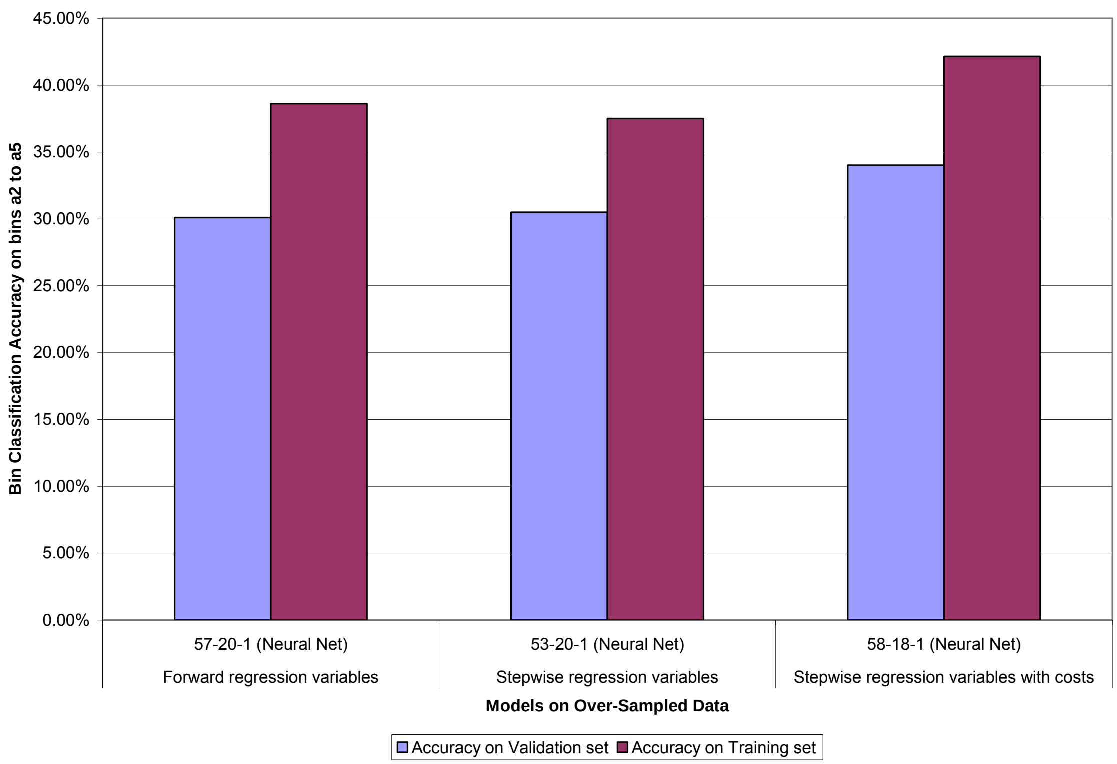


Figure 9: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on Over-sampled data)

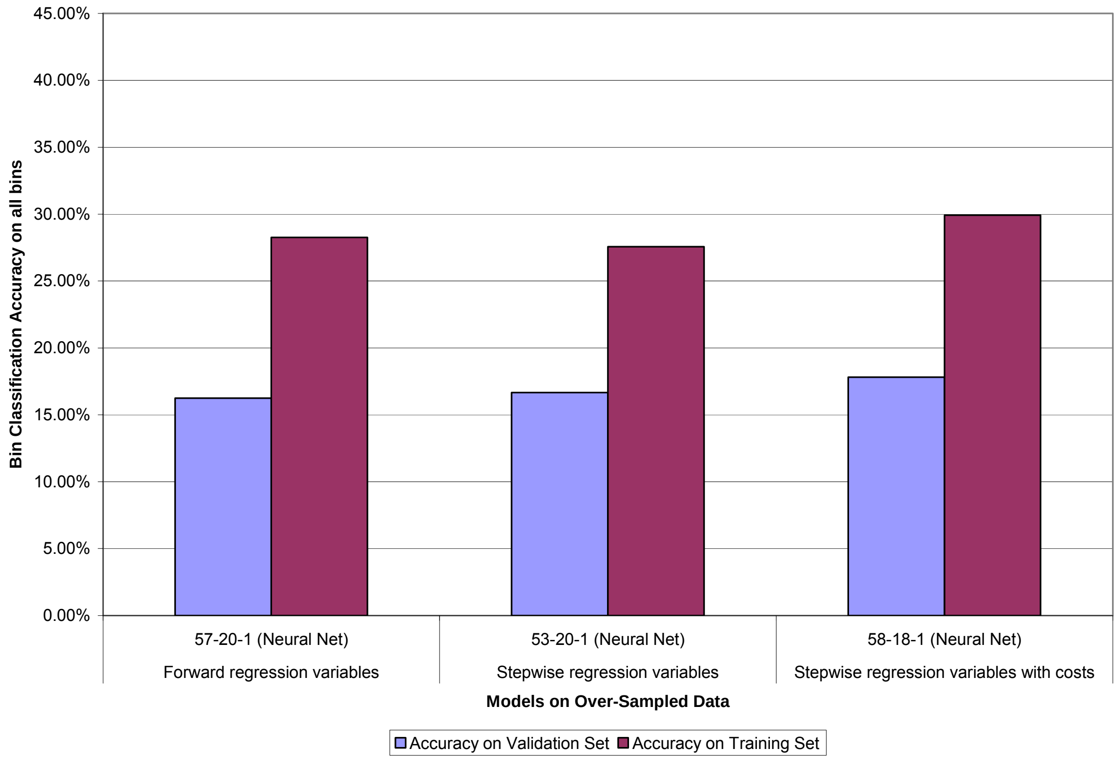


Figure 10: Bin Classification Accuracy on Full Range of change in VTR (all bins) for validation & training set (Different models on Over-sampled data)

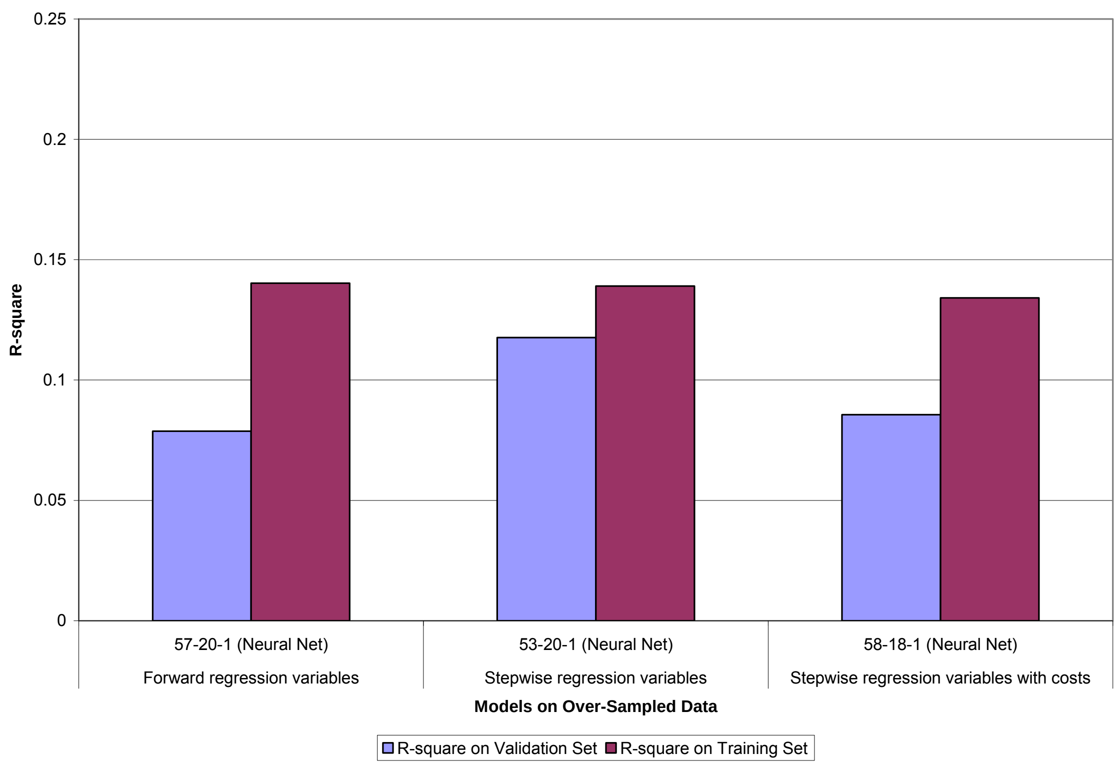


Figure 11: R-square for validation & training set (Different models on Over-sampled data)

As can be seen from the figures, for all the models the 'bin classification accuracy on moderate range of change in VTR' has been greatly improved, with little change in total accuracy. However, the R-square value on the validation set has gone down drastically. The model which got the best of three measures was the one constructed with stepwise regression attributes.

But it can be seen from the scatter plot of this model in Figure 12, that for many of the records with positive change in VTR, the model predicted negative change in VTR. The best model should predict both negative as well as positive changes in VTR with a high degree of accuracy. However, this particular model was biased towards predicting negative changes in VTR, and hence fails to achieve the required goal.

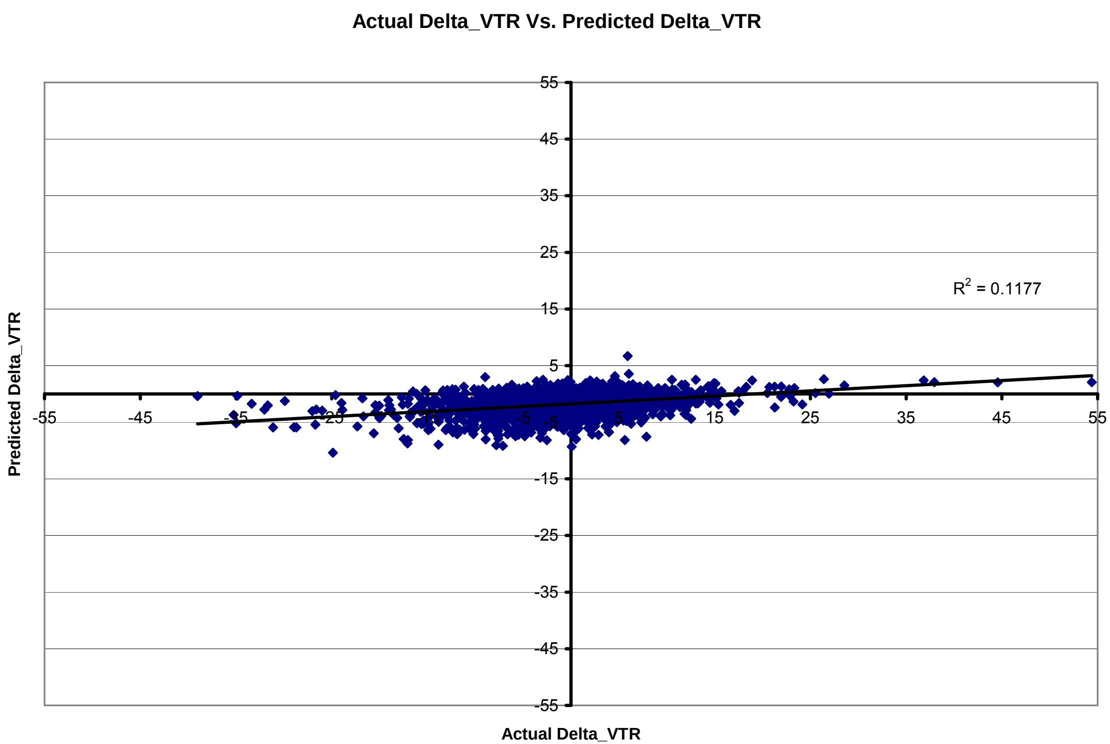


Figure 12: Scatter plot for stepwise linear regression model

Phase III: Los Angeles Data - records with no incentives removed

Some of the cases in the data sets did not have any incentives implemented. Such records would not be helpful in trying to find the effectiveness of incentives on the reduction in vehicle trip rate and were removed from the two datasets respectively, with 18,140 records remaining in the final data set.

Table 8: Los Angeles data with records with no Incentives Removed – No. of Records in bins

range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)							
		> - 7	[- 7 to - 4)	[- 4 to - 2)	[- 2 to -0)	[-0 to 1)	[1 to 2.5)	[2.5 to 5)	5 >=
Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**
Validation	**1838**	292	220	261	290	173	157	186	259
Training	**16302**	2543	2310	2192	2635	1414	1479	1714	2015

Several models were built on the data with these records, but with different variable sets.

For some models, the costs associated with each incentive were added into the dataset. Because the model with stepwise regression variables gave the best results of the full sample data, it was once again used on this data. The variable set selected by the model is shown in Table 9.

Table 9: Variables stepwise regression--data with associated incentive costs

Variable	Description
VTR_CUTR	Vehicle trip rate
TRANSIT	Share employees commuting using transit
BMM	Posted Materials (Marketing)
TELECOMMUTE	Share employees tele-commuting
CWW440	Share of 4/40 CWW days off
BIKE	Share employees commuting using bike
ISS	Cafeteria, ATM's, Postal, Fitness Center
DW4	4/40 Compressed Work Week Schedule
BGT_DOLLAR_AMT	Taxi Guaranteed Return Trip amount
CWW336	Share of employees on 3/36 CWW
TARGET_AVR	Target AVR required (3 zones)
BGE	Emergencies Guaranteed Return Trip
DNT	Additional Time Off with Pay
BMO	Other Marketing Elements
BUS	Share employees commuting in bus
CAR2	Share employees commuting two together
DFT_DOLLAR_AMT	On-Going Transit Subsidies amount
RS_MATCH	Combined ride-share match
BMS	Special Interest Club (Biking, Walking)
BGC	Company Vehicle Guaranteed Return Trip
BGR_DOLLAR_AMT	Rental Car Guaranteed Return Trip amount
BMR	Company Recognition (Marketing)
DPC	Increased Parking Costs for Drive Alones
DPO	Other Parking Management Strategies
IST_DOLLAR_AMT	Transit Information or Pass Sales amount
BGT	Taxi Guaranteed Return Trip
DW4_DOLLAR_AMT	4/40 Compressed Work Week Schedule amount
BGO	Other Guaranteed Return Trip Program
BFR	Bike Racks and Bike Lockers
BFO	Other Facility Improvements
BGU	Unscheduled Overtime Guaranteed Return
CAR4	Share employees commuting four together

Variable	Description
BMN	New Hire Orientation (Marketing)
DWO	Other Compressed Work Week Schedule
COMPRESSED	Combined compressed WW
CWW980	Share of 9/80 CWW days off

A neural network model was built on the dataset containing the variables selected by the above regression model. Also a neural network model was built on variables selected by the stepwise regression on the full sample data with respective incentive costs added.

To reduce the complexity of the model, all of the individual incentives were removed and a simple model was built on the data containing only the worksite characteristics like mode-splits and the grouped incentives.

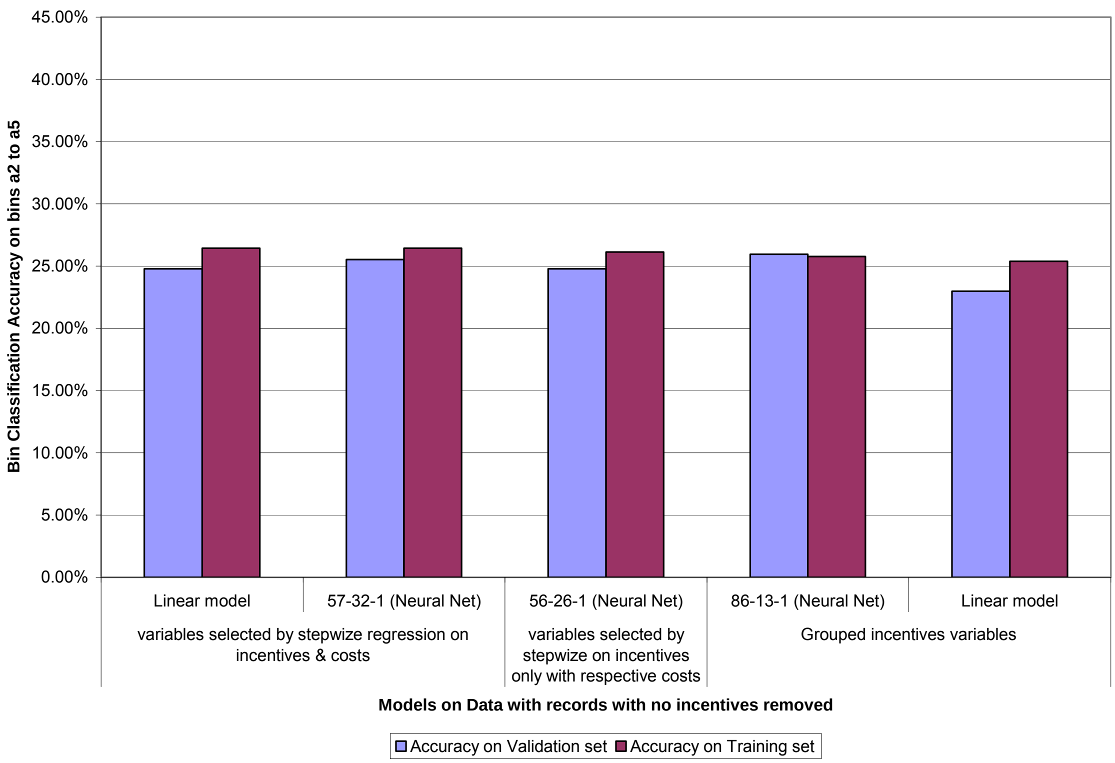


Figure 13: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on data – records with no incentives removed)

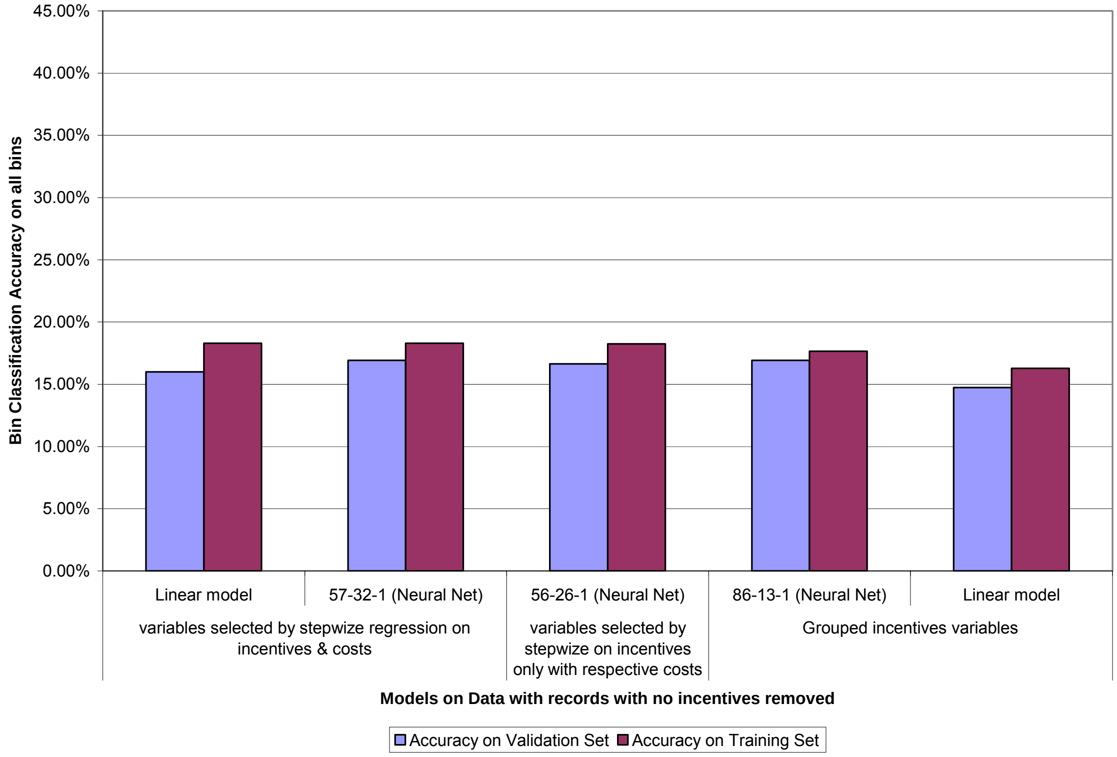


Figure 14: Bin Classification Accuracy on Full Range of change in VTR (all bins) for validation & training set (Different models on data – records with no incentives removed)

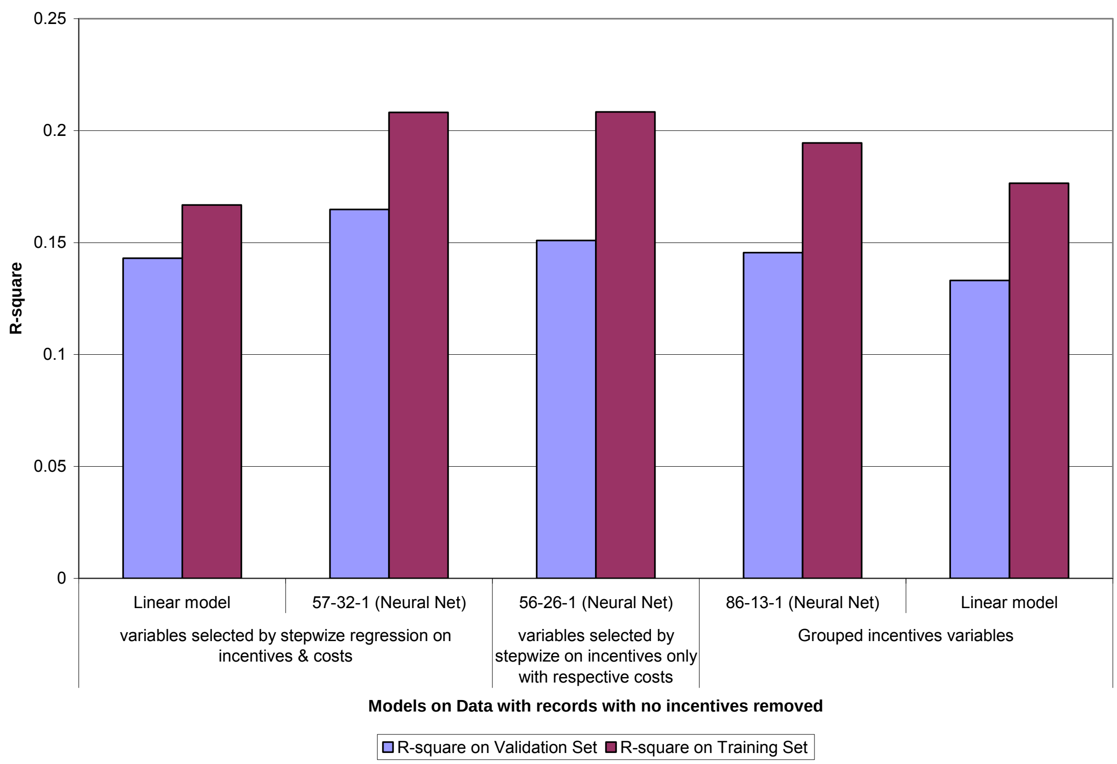


Figure 15: R-square for validation & training set (Different models on data – records with no incentives removed)

It can be seen from Figures 13 and 14, that the two neural network models, one built using the stepwise regression variables on incentives and costs (M1) and the other built on grouped incentives variables (M2) were two frontrunners for the best model. The bin accuracies for this model are shown in Table 10.

Table 10: Detailed accuracies on bins

	range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
			> - 7	[- 7 to - 4)	[- 4 to - 2)	[- 2 to -0)	[-0 to 1)	[1 to 2.5)	[2.5 to 5)	5 >=		
	Bin Number		**a1**	**a2**	**a3**	**A4**	**a5**	**a6**	**a7**	**a8**		
	Validation	**1838**	292	220	261	290	173	157	186	259	Weighted	R-square
	Training	**16302**	2543	2310	2192	2635	1414	1479	1714	2015	Avg on	
											a2 to a5	
M1	**Exact Validation**	16.92%	7.19%	19.55%	27.59%	35.52%	13.29%	12.10%	6.99%	6.56%	25.53%	0.1649
	Exact Training	18.30%	7.75%	22.68%	30.34%	33.02%	14.29%	12.64%	9.33%	8.83%		0.2081
	One-off Validation	47.82%	31.85%	54.09%	77.78%	67.24%	56.65%	36.94%	27.42%	23.94%	65.15%	
	One-Off Training	49.53%	37.48%	54.37%	76.05%	68.92%	58.77%	35.36%	27.19%	27.89%		
M2	**Exact Validation**	16.92%	4.11%	19.09%	27.59%	36.21%	15.03%	11.46%	9.14%	7.34%	25.95%	0.1456
	Exact Training	17.65%	5.82%	20.13%	30.93%	34.04%	11.60%	12.04%	9.63%	9.03%		0.1945
	One-off Validation	45.48%	27.40%	50.91%	75.48%	71.72%	51.45%	32.48%	23.12%	21.62%	64.19%	
	One-Off Training	48.77%	31.97%	52.03%	78.88%	72.87%	57.99%	31.78%	25.73%	27.54%		

It can be seen from table 10, that both the models have same 'bin classification accuracy on full range of change in VTR'. The grouped incentive model has better 'bin classification accuracy on moderate range of change in VTR' than the other model with less R-square value. It can also be seen that the grouped incentive model was able to get better bin accuracies on positive range of change in VTR.

The scatter plots for both models on the validation set are shown in Figures 16 and 17.

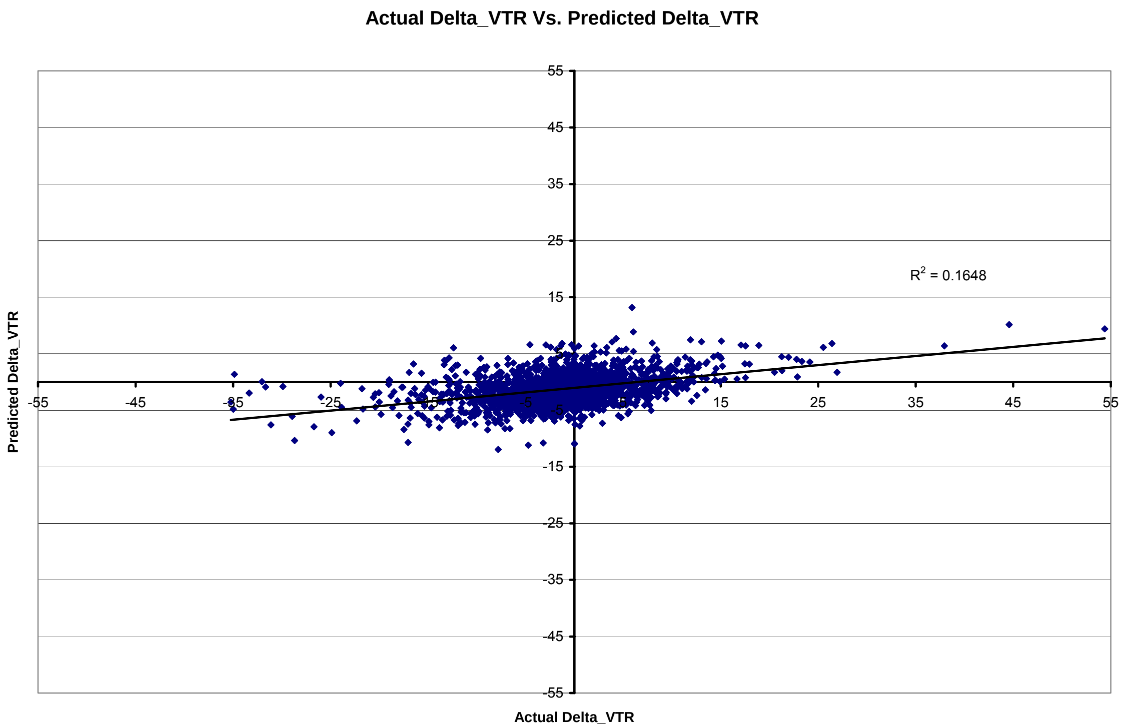


Figure 16 : Scatter plot for NN model built on stepwise regression variable

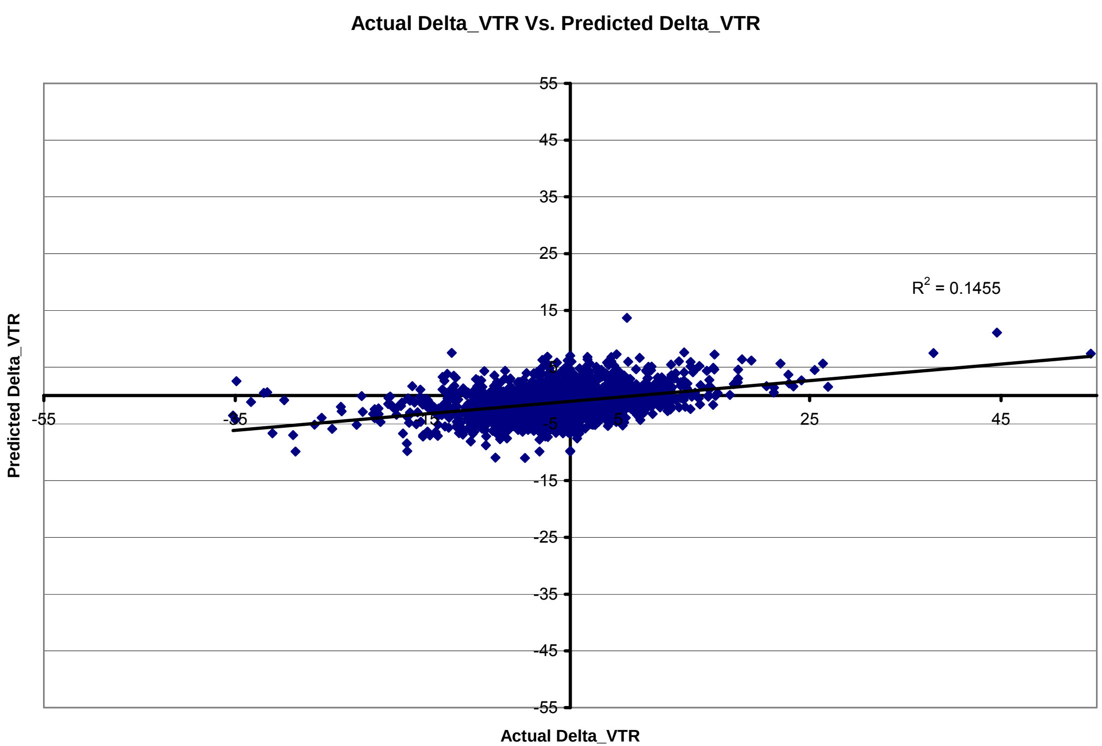


Figure 17: Scatter plot for NN model built on Grouped incentive variable

Phase IV: Los Angeles Data - records with no incentives & outliers removed

The dataset contained some records with very large changes in the VTR in both directions. It was felt that such large changes might not be attributed to the incentives plans implemented by the companies but rather to other non recorded measures. So the records with change in VTR less than –15.2 and greater than 12.5 were considered as outliers and were removed from the training and validation data. The reason for the selection of these cut-off values for change in VTR was that, these values that were outside of the normal distribution of the dataset. In order to obtain these values, SPSS 11.1 was used to identify values in the dataset which were greater or less than three standard deviations of the mean. This resulted in 790 records being dropped from the validation set and 7,432 dropped from the training set.

Table 11: Los Angeles data with records no incentives & outliers removed- – No. of Records in bins

range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)							
		[- 15.2 to - 7)	[- 7 to - 4)	[- 4 to - 2)	[- 2 to -0)	[-0 to 1)	[1 to 2.5)	[2.5 to 5)	[5 to 12.5)
Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**
Validation	**1738**	236	220	261	290	173	157	186	215
Training	**15490**	2103	2310	2192	2635	1414	1479	1714	1643

Three models were built again on this data on the variables selected by the models which had given good accuracies measures

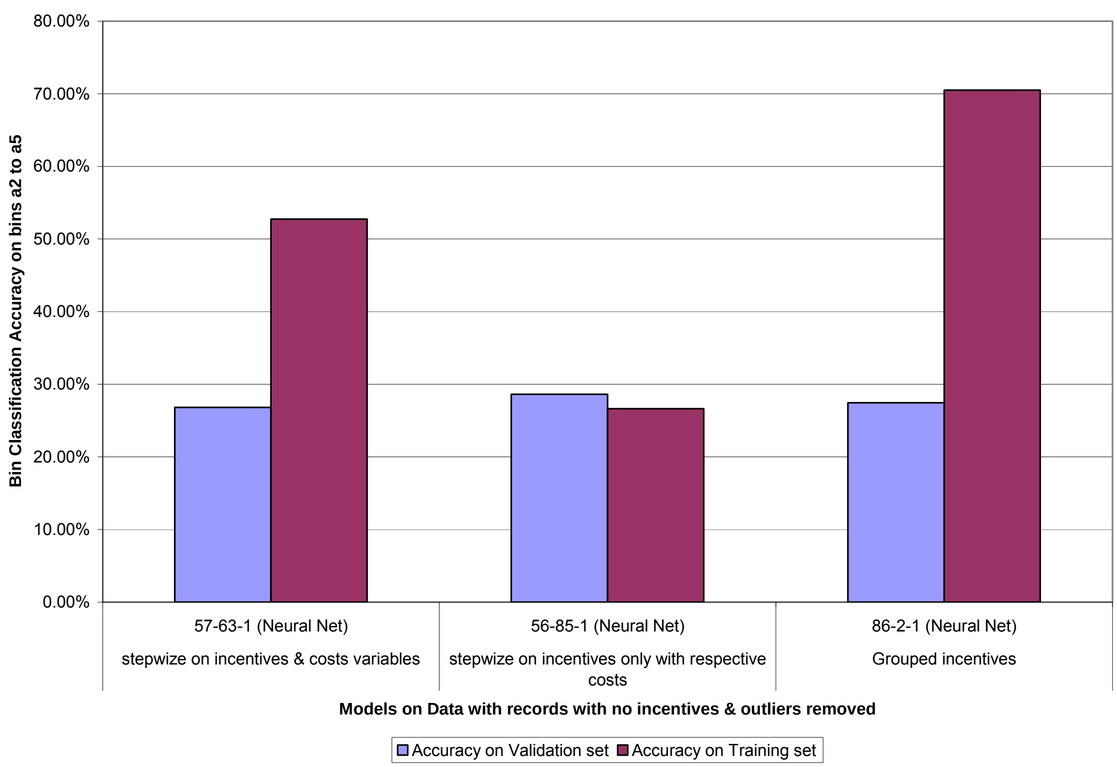


Figure 18: Bin Classification Accuracy on Moderate Range of change in VTR(a2 to a5) for validation & training set (Different models on data – records with no incentives & outliers removed)

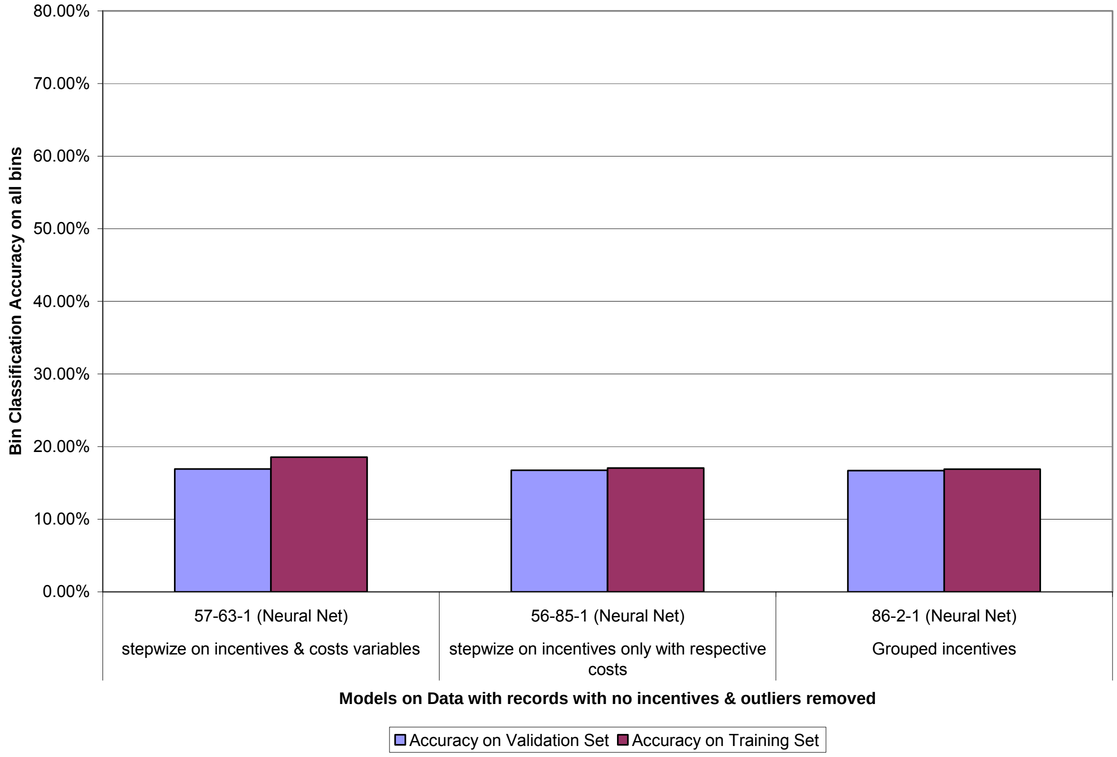


Figure 19: Bin Classification Accuracy on Full Range of change in VTR(all bins) for validation & training set (Different models on data – records with no incentives & outliers removed)

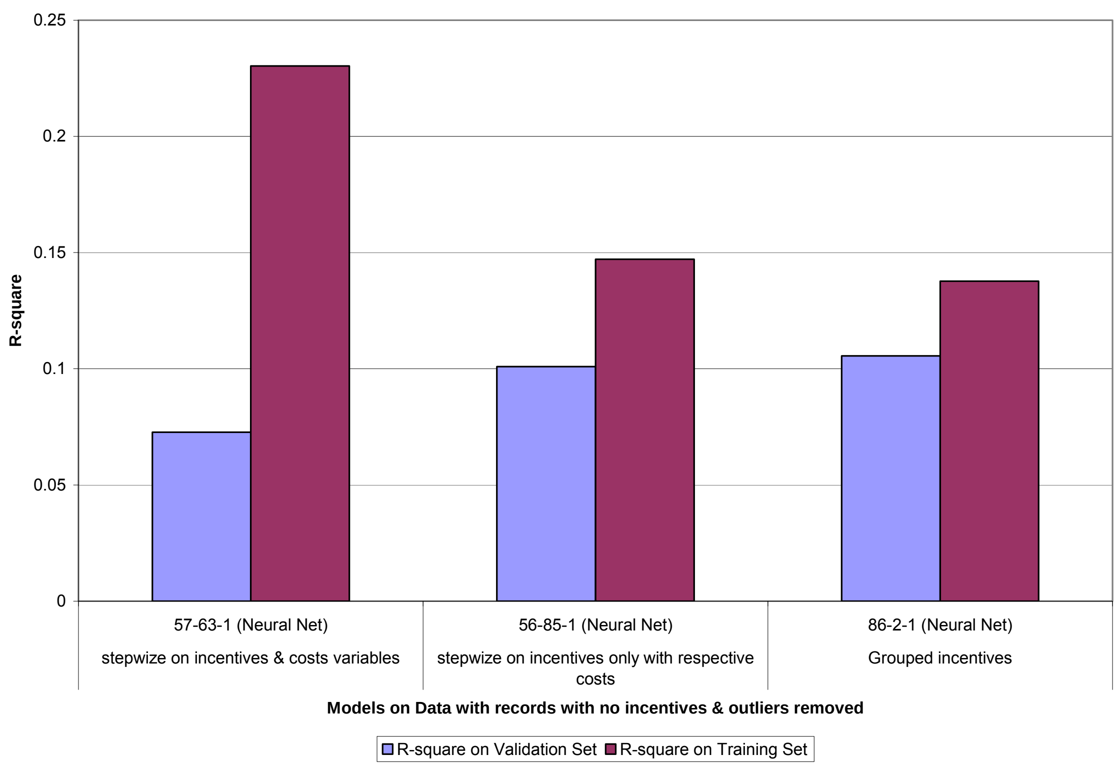


Figure 20: R-square for validation & training set (Different models on data - records with no incentives & outliers removed)

Out of all of the models built without the outlier data, the model that got the best of three measures was the one built from stepwise regression on incentives with respective costs added. Figure 21 shows the scatter plot for this model on the validation set and it is apparent that this model was not a good predictor of change in VTR.

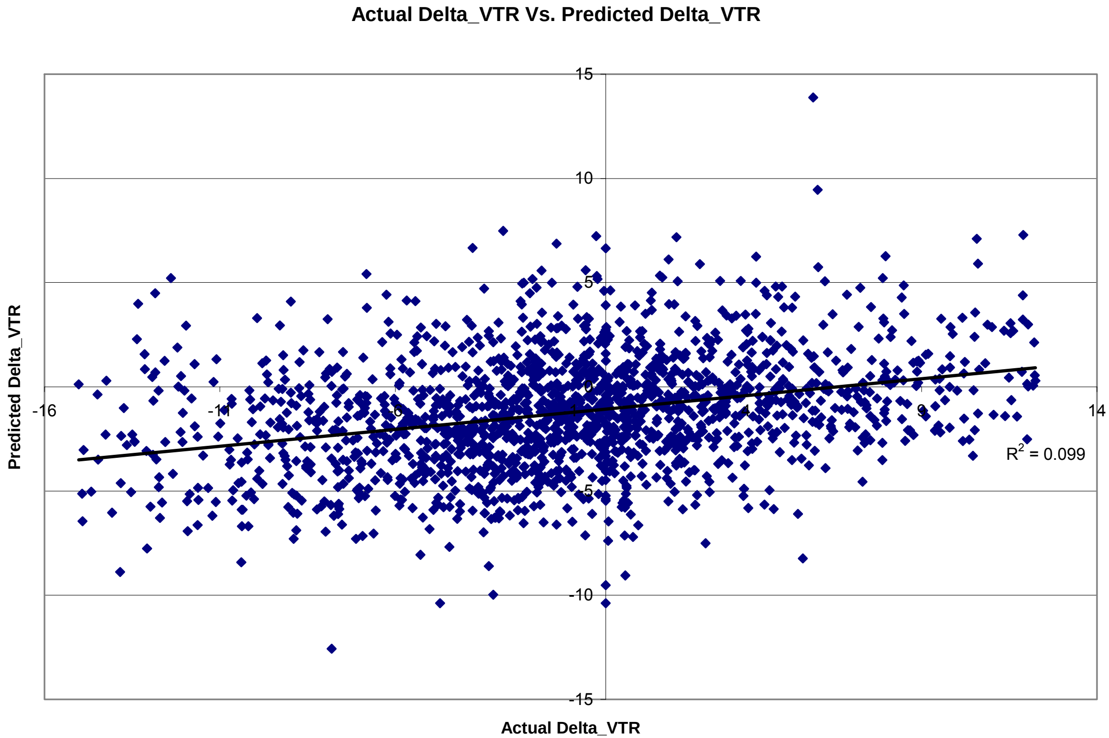


Figure 21: Scatter plot for model built on stepwise regression on incentives with respective costs added

Though the assumption about outliers might have been true, the removal of the outlier records further reduced the prediction power of these models.

Recommended Model

In phase I and II of model building, the datasets contained some records which had no information about the incentives each worksite had implemented. This might have reduced the importance of incentives in predicting the change in VTR. In phase III, data was cleaned by deleting the records that did not have any information regarding implemented incentives. There was a considerable improvement in the accuracies in all the models as compared to the models built in the previous two phases.

In phase IV, although the outlier records were removed from the data, the total accuracies and the R-square values of the models went down. The two best candidate models were the ones built using the stepwise regression variables on incentives and costs and the other built on grouped incentives.

The reasons the model built on stepwise regression variables on incentives and costs was considered as candidate was (Table 10)

1. It has 'bin classification accuracy on moderate range of change in VTR' of 25.53%
2. It has 'bin classification accuracy on full range of change in VTR' of 16.92% which is better than the random choice accuracy of 12.5%
3. It has the best 'R-square' value of 0.1649

The reasons the model built on grouped incentives data was considered as candidate was (Table 10)

1. It has 'bin classification accuracy on moderate range of change in VTR' of 25.95%
2. It has 'bin classification accuracy on full range of change in VTR' of 16.92% which is better than the random choice accuracy of 12.5%
3. It has a 'R-square' value of 0.1456
4. It is built on simple grouped incentive variable set

The variable set selected by the regression model was not complete as it did not cover all of the incentives, did contain some very detailed specific incentives like Taxi Guaranteed Return Trip dollar amount, etc. which the user of the model might not be able to provide. Thus due to the above reasons, the variable set would not have appealed to the transportation planners and engineers when compared to the simple grouped incentives variable set. So considering these facts, the recommended model is the one built on the grouped incentives with variables shown in the Table 12.

Table 12: Variables for recommended model Los Angeles Model

Variable	Description
T_AVR1.3	down town zone
T_AVR1.5	suburbs zone
T_AVR1.75	outside suburb zone
VTR_CUTR	Vehicle trip rate
CAR1	Share of employees commuting by motorcycle
Motorcycle	Share of employees commuting alone in car
CAR2	Share of employees commuting two together
CAR3	Share of employees commuting three together
CAR4	Share of employees commuting four together
CAR5	Share of employees commuting five together
CAR6	Share of employees commuting six together
VAN_CUTR	Share of employees commuting in van
BUS	Share of employees commuting in bus
TRANSIT	Share of employees commuting using transit
WALK	Share of employees commuting walking
BIKE	Share of employees commuting using bike
TELECOMMUTE	Share of employees telecommuting
CWW336	Share of 3/36 CWW days off
CWW440	Share of 4/40 CWW days off
CWW980	Share of 9/80 CWW days off
FACILITY_AMENITIES	Passenger Loading Areas, Other Facility Improvements, Preferential Parking Areas, Bike Racks and Bike Lockers, Shower and Lockers
RIDE_HOME	TMA/TMO Provided Guaranteed Return Trip, Company Vehicle Guaranteed Return Trip, Emergencies Guaranteed Return Trip, Other Guaranteed Return Trip Program, Rental Car Guaranteed Return Trip, Taxi Guaranteed Return Trip, Unscheduled Overtime Guaranteed Return
FLEXTIME	Flextime for Ride sharers (Work Shifts), Flextime for Ride sharers (Grace Period)
MARKETING	Commuter Information Center, Commuter Fairs (Marketing), Focus Groups (Marketing), Posted Materials (Marketing), New Hire Orientation (Marketing), Other Marketing Elements, Personal Communication (Marketing), Company Recognition (Marketing), Special Interest Club (Biking, Walking), TMA/TMO Membership (Marketing), Written Materials (Marketing), Zip Code Meetings (Marketing)
RS_MATCH	Regional Commuter Management Agency, Employer-Based Rideshare Matching System
FINANCIAL	Transportation Allowances, On-Going Bike-to-Work Subsidies, On-Going Carpooling Subsidies, Other Direct Financial Subsidies, On-Going Walk-to-Work Subsidies
PARKMGT$	Increased Parking Costs for Drive Alones, Other Parking Management Strategies, Subsidized Parking for Ride sharers
TELE_GRP	Work at Home (Telecommuting), Work at Satellite Center (Telecommuting)
COMPRESSED	3/36 Compressed Work Week Schedule, 4/40 Compressed Work Week Schedule, 9/80 Compressed Work Week Schedule, Other Compressed Work Week Schedule
VANPOOL	Company Owned/Leased Vanpools
ONSITE	On-Site Childcare Service, Other On-Site Services, Cafeteria, ATM's, Postal, Fitness Center, Transit Information or Pass Sales
DIRECT_NONFINANCIAL	Auto Services (Fuel, Oil, Tune-Up), Gift Certificates, Free Meals, Other Direct Non-Financial Incentives, Catalogue Points, Additional Time Off with Pay, Drawings, Free Meals, Certificates, etc.
OTHER	Other Not Classified by Other Codes
COMMTAX	Introductory Transit Passes or Subsidies, Subsidized Vanpool Seats, On-Going Transit Subsidies, On-Going Vanpooling Subsidies
DeltaVTR	Change in VTR

TUCSON MODELS

Phase I: Tucson Full Sample Data

The Tucson dataset consisted of 1,121 total examples. Approximately ten percent of the dataset was left aside as validation set while the rest became training/testing set. Because the training/testing set contained just 1,009 examples, ten percent of the random data of was set aside as the testing set. The eight bin ranges and the number of examples in each bin for validation and training/testing set are shown in Table 13.

Table 13: Tucson Full Sample Data – No. of Records in bins

Range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)							
		> - 7	[- 7 to - 4.5)	[- 4.5 to – 3)	[- 3 to -1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=
Bin Number		a1	a2	a3	a4	a5	a6	A7	a8
Validation	112	4	6	6	13	19	25	22	17
Training	1009	37	52	56	116	171	222	200	155

Neural network models were built on this data with different variable selection parameters. All types of regression models were also built. Figures 22, 23, and 24, show the comparison of the different performance measures on this data.

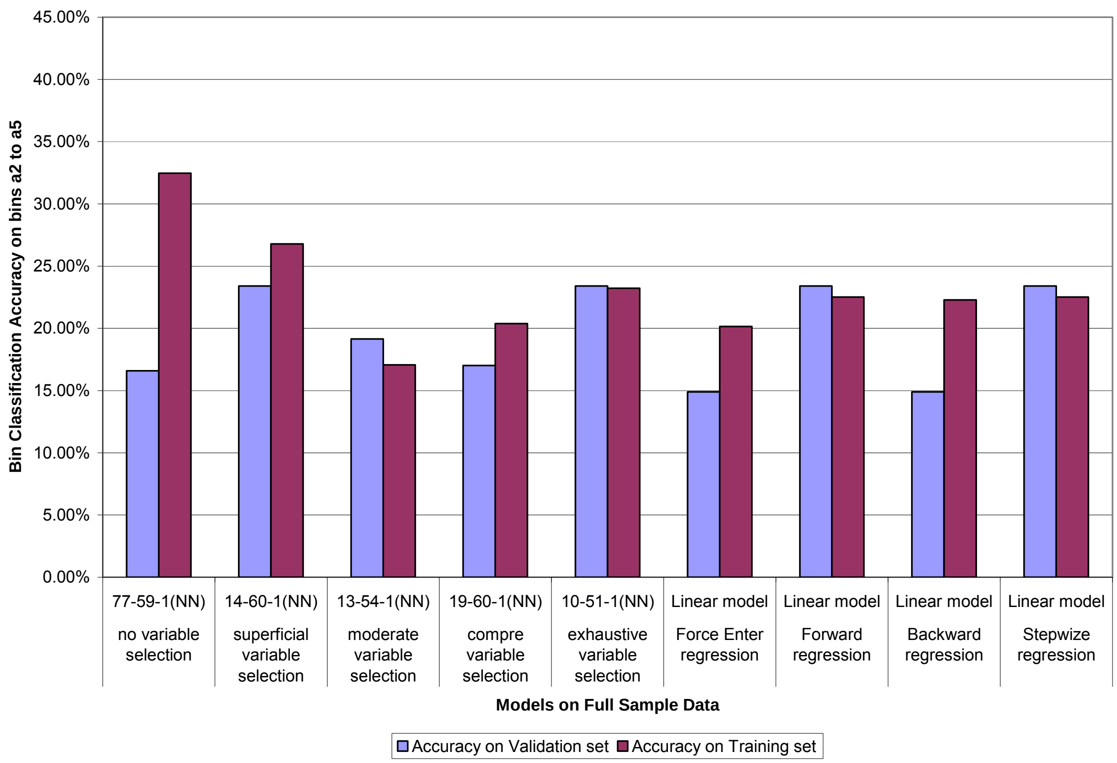


Figure 22: Bin Classification Accuracy on Moderate Range of change in VTR(a2 to a5) for validation & training set (Different models on Full sample data)

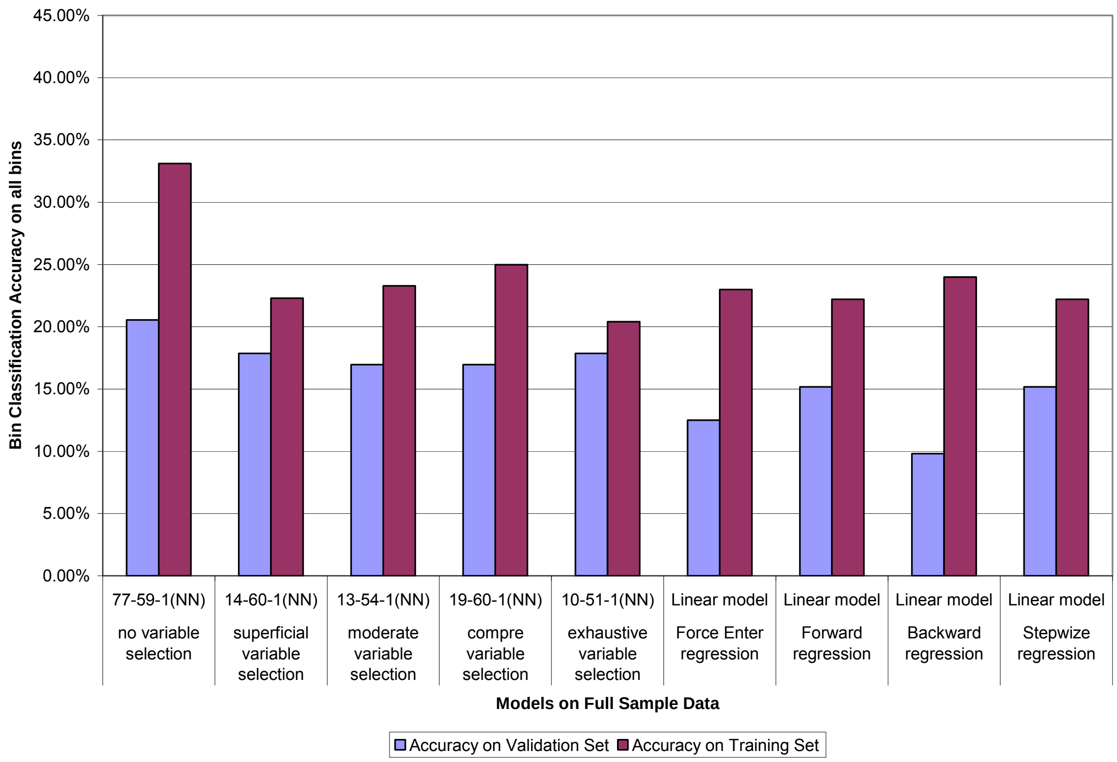


Figure 23: Bin Classification Accuracy (all bins) (Full sample data)

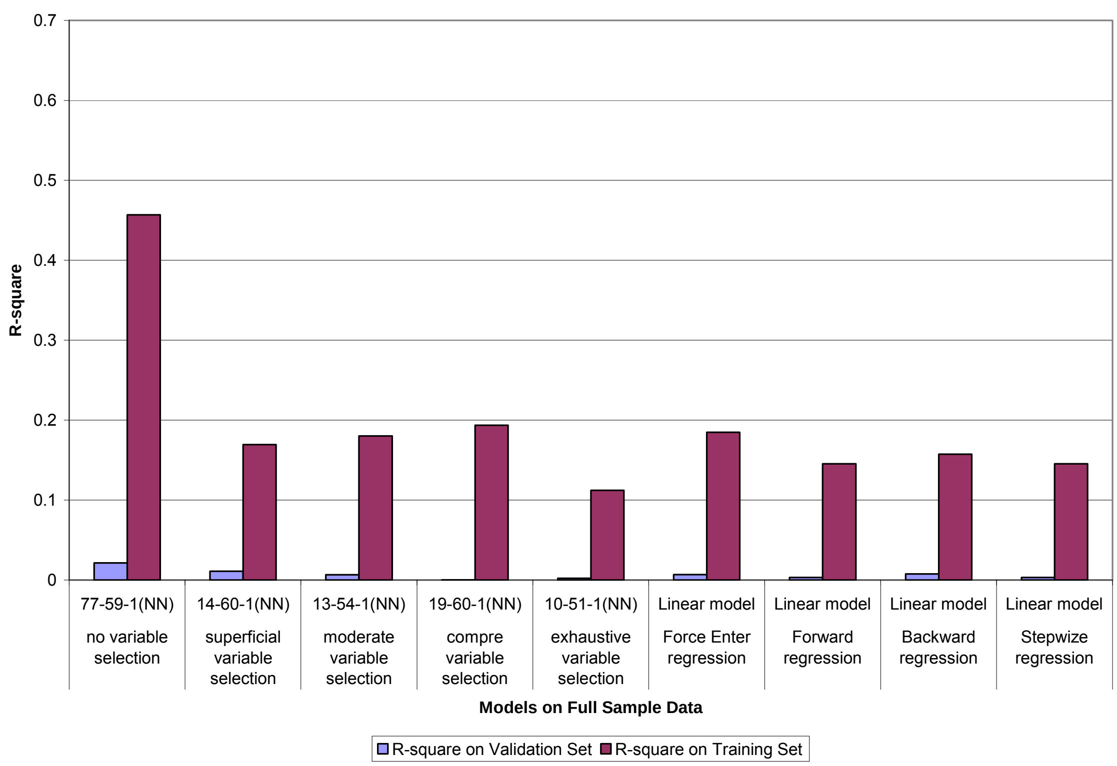


Figure 24: R-square for validation & training set (Different models on Full sample data)

The variables selected by these models are shown below in Table 14.

Table 14: Variables selected by different models on Full Sample data

variables	Predicts variable selection					Regression Approaches			
	no var	Superficial	moderate	Comprehensive	Exhaustive	forced enter	For ward	Backward	Stepwise
Alone share	√		√	√					
Bus share	√	√	√	√	√	√		√	
Carpool + Vanpool share	√	√			√	√		√	
walk share	√	√	√	√		√		√	
Motorcycle share	√	√	√	√		√	√	√	√
Bicycle share	√	√	√		√	√	√	√	√
Average Miles traveled	√		√	√	√	√			
Average minutes traveled	√			√	√	√			
3/36 compressed work week share	√		√	√		√			
4/40 compressed work week share	√	√	√	√		√	√	√	√
8/80 compressed work week share	√	√	√			√			
Number of Employees	√	√				√			
Adjusted Work Hours									
Alternative Fuel Vehicles	√					√			
Alternate Mode Campaign	√			√		√			
Alternate Mode Information	√					√			
Bicycle Campaign	√			√		√			
Bicycle Racks	√					√			
Bus Pass Sales on Site						√			
Bus Subsidy						√			
Busing Vehicle	√					√			
Carpool Campaign	√					√			
Carpool Subsidy	√					√			
Carpooling Vehicle	√					√			
Compressed Work Week	√					√			
Coordination with Transit Provider	√					√			
Covered Parking	√	√				√			
Daycare Facilities on Site	√					√			
Dissemination of Air Quality Information	√				√	√			
Drawing for Prizes	√	√	√	√		√			
Employee Shifts Between Sites	√				√	√		√	
Fee for Parking	√					√			
Field Worker	√					√			
Guaranteed Ride Home Program	√				√	√			
Incentive Programs	√					√			
Incentives for Employees to Live Close	√					√			
Information Center	√					√			
Map Board	√			√		√			
Matching Service									
New Employee Information	√					√			
Newsletter Articles	√		√			√			
Post Air Quality Information	√			√		√			
Preferred Parking									
Rebate not to Use Parking	√					√			
Rideshare Committee	√			√		√			
Showers/Lockers	√					√			
Shuttle Vehicle	√			√		√		√	
Speakers	√					√		√	
Staging Area	√		√			√			
Transportation Fair						√			
Vanpooling Vehicle						√			
Vanpool Subsidy	√	√				√	√	√	√
Vanpooling Vehicle	√					√			
Walking Campaign	√					√			
Work at Home	√					√			
VTR	√		√			√	√	√	√

It can be seen from the performance measures charts, that neural network models built using superficial and exhaustive variable selection and linear models built using forward and stepwise regression were able to get the best 'bin classification accuracy on moderate range of change in VTR' (23.40 percent). But, all of these models had very poor R-square values. It can be seen from the variable selection table, that the variables selected by all of these models contain very few incentive plans, thus defeating the purpose of the models to predict change in VTR using many incentives. All the neural network and regression models built using variable selection selected very few incentives variables and were considered as unsuitable models.

The neural network model built with no variable selection got the best 'bin classification accuracy on full range of change in VTR' (20.54 percent) and best R-square value (0.022), and could be considered as a better model. The detailed bin accuracies are shown in Table 15 and the scatter plot in Figure 25.

Table 15: Detailed accuracies for neural net model with no variable selection

range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
		> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to - 1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=		
Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**		
Validation	**112**	4	6	6	13	19	25	22	17	Weighted	R-square
Training	**1009**	37	52	56	116	171	222	200	155	Avg on	
										a2 to a5	
Exact Validation	20.54%	0.00%	0.00%	20.00%	23.53%	11.76%	14.29%	36.84%	25.00%	16.60%	0.022
Exact Training	33.10%	0.00%	24.56%	29.85%	28.23%	39.08%	27.67%	31.49%	50.61%	32.46%	0.457
One-off Validation	43.75%	0.00%	0.00%	60.00%	41.18%	41.18%	52.38%	47.37%	45.00%	41.28%	
One-Off Training	70.86%	38.89%	57.89%	49.25%	65.32%	79.31%	76.21%	71.82%	78.66%		

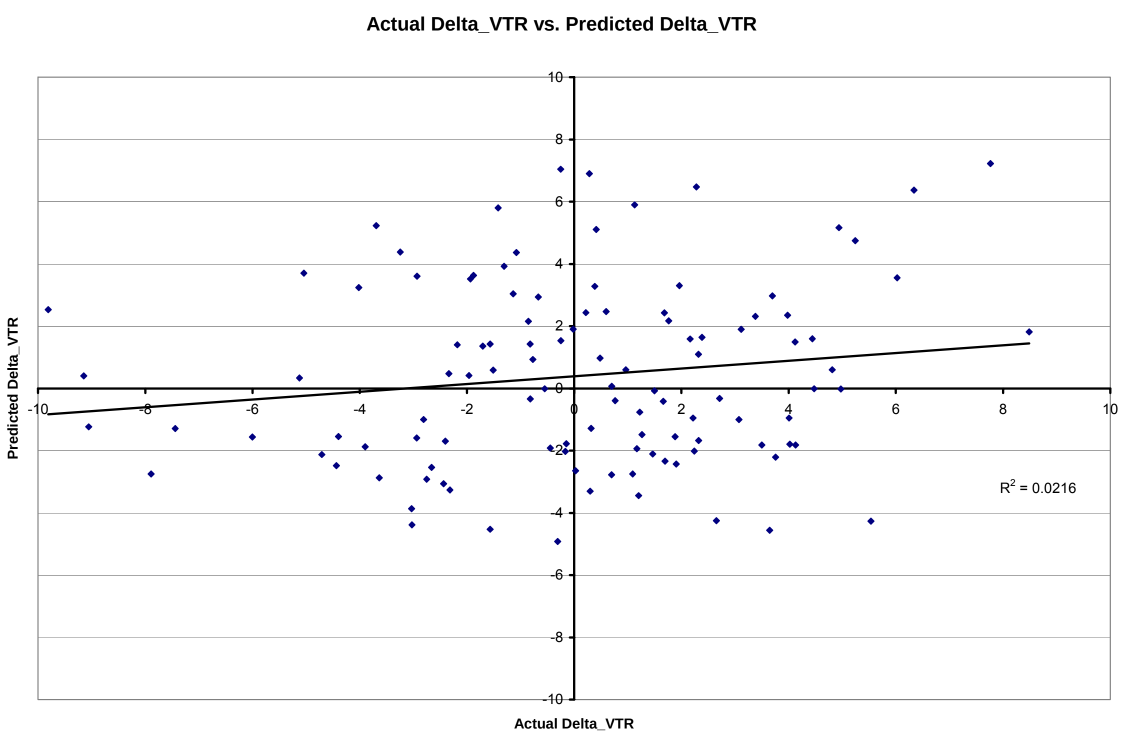


Figure 25: Scatter plot for neural net model with no variable selection

To reduce to complexity of the models, all the individual incentives were removed and then replaced with the grouped incentives and simple neural network with no variable selection and forced enter linear regression models were built on the data just containing the grouped incentives and worksite characteristics like mode-splits. The variables for the Tucson grouped incentives model are shown in Table 16.

Table 16: Variables with grouped incentives

Variable	Description
AloneShare	Alone share
BusShare	Bus share
CVpoolShare	Carpool + Vanpool share
WalkShare	walk share
McycleShare	Motorcycle share
BcycleShare	Bicycle share
HiMiles	Average Miles traveled
HiMinute	Average minutes traveled
CWW336	3/36 compressed work week share
CWW440	4/40 compressed work week share
CWW980	8/80 compressed work week share
NoEmp	No. of employees
FACILITY_AMENITIES	facilities & amenities (grouped incentives)
GRH	Guaranteed ride home programs (grouped incentives)
FLEX	flexible timing (grouped incentives)
Mrkt	Marketing programs(grouped incentives)
RS_MATCH	Ride share matching programs(grouped incentives)
FINANCIAL	financial incentives(grouped incentives)
PARKMGT	Parking management (grouped incentives)
TELE	Telecommute program (grouped incentives)
CWW	Compressed work week program(grouped incentives)
VANPOOL	Vanpool vehicles (grouped incentives)
onsite	onsite incentives (grouped incentives)
direct_nonfinan	Non financial incentives (grouped incentives)
commtax	commuter tax benefit incentives (grouped incentives)
VTR	Vehicle trip rate

Figures 26, 27, and 28 show the comparison of the different performance measures on this grouped incentive data.

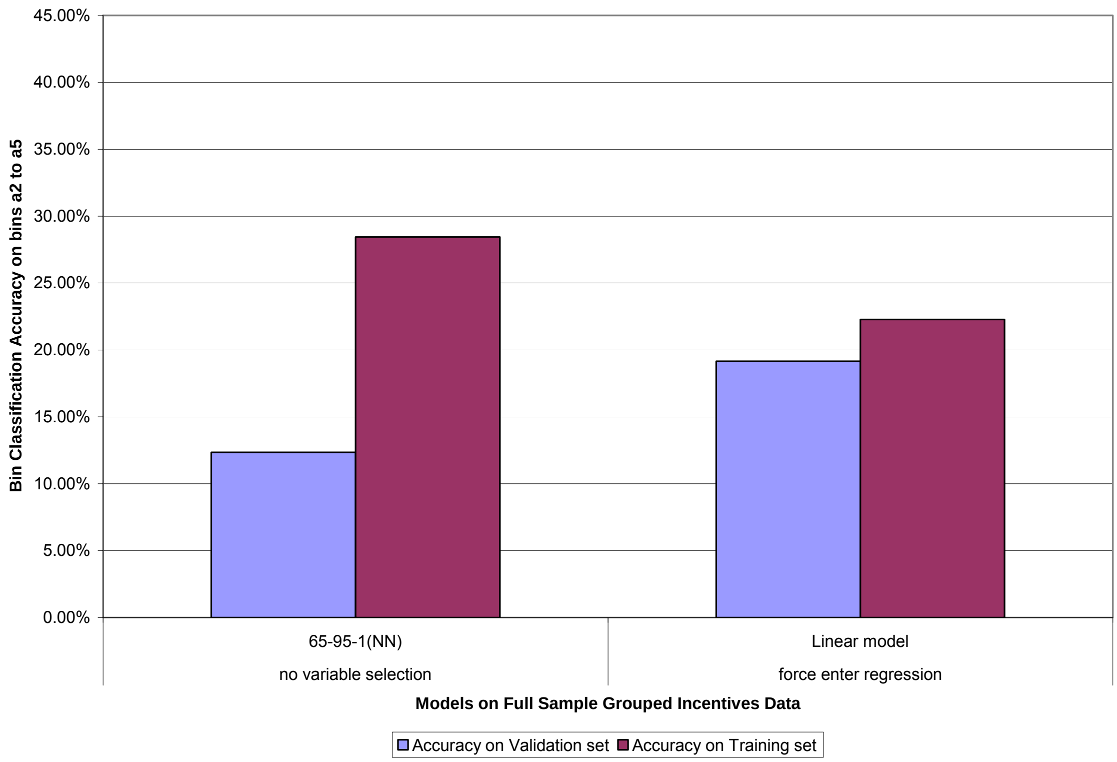


Figure 26: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) (Different models on Full sample Grouped Incentive data)

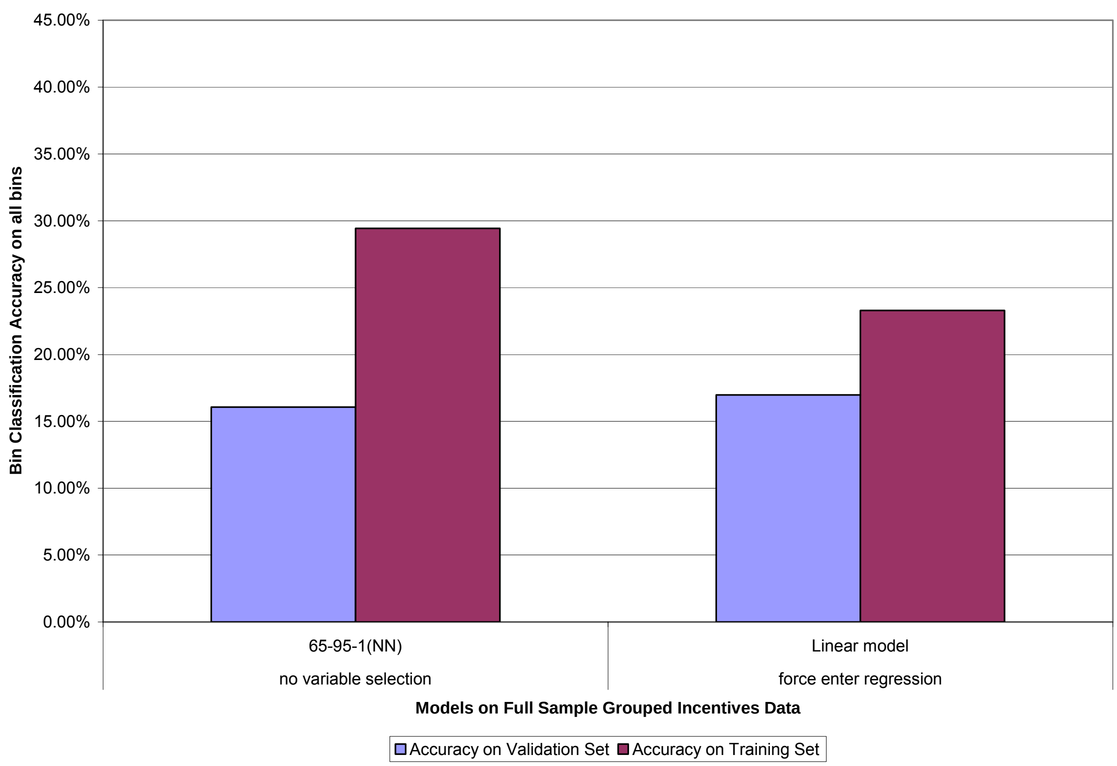


Figure 27: Bin Classification Accuracy on Full Range of change in VTR (all bins) (Different models on Full sample Grouped Incentive data)

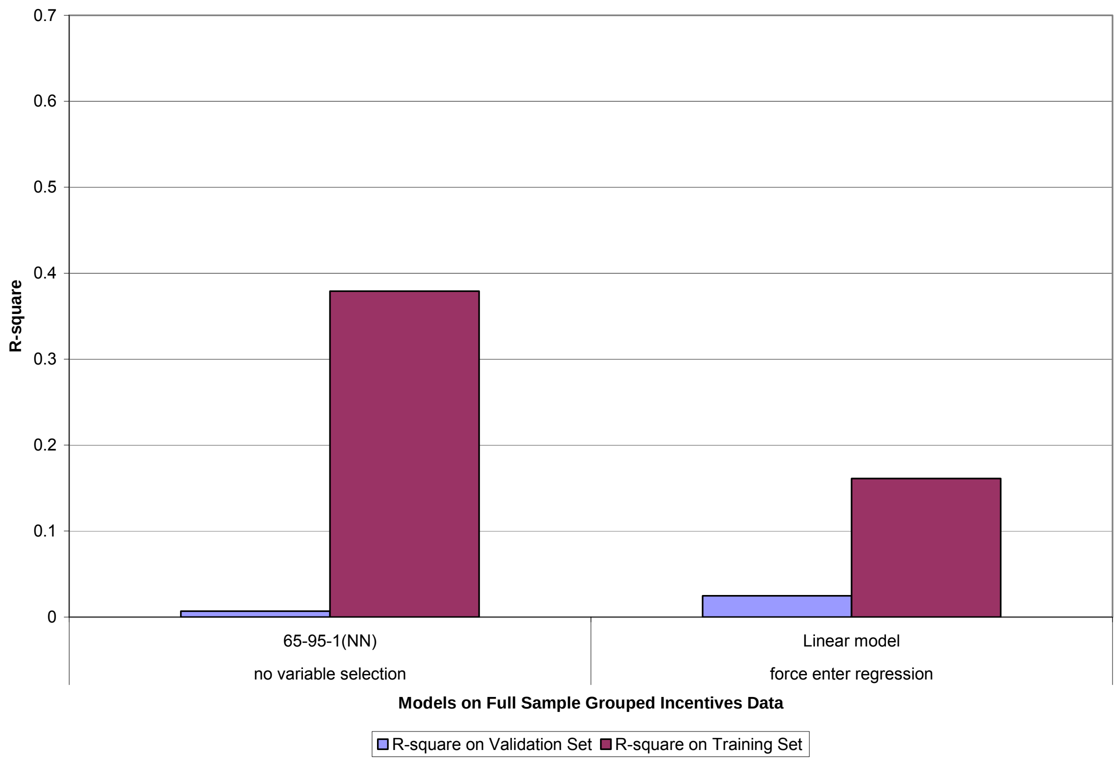


Figure 28: R-square for validation & training set (Different models on Full sample Grouped Incentive data)

It can be seen from Figures 26 and 27 that the neural network gets much better accuracy on the training set than the regression models but lower accuracy on the validation set. A possible explanation for this anomalous behavior is that the neural network model might be over-fitting the training data, thereby making it useless for the validation set. In these two models, the forced enter regression model was clearly the best. The detailed accuracies of the bins for this model are shown in Table 17.

Table 17: Detailed accuracies for forced enter regression model

range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
		> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to - 1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=		
Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**		
Validation	112	4	6	6	13	19	25	22	17	Weighted	R-square
Training	1009	37	52	56	116	171	222	200	155	Avg on a2 to a5	
Exact Validation	16.96%	0.00%	0.00%	0.00%	17.65%	35.29%	33.33%	10.53%	5.00%	19.15%	0.0248
Exact Training	23.29%	0.00%	0.00%	0.00%	9.68%	47.13%	36.41%	16.02%	22.56%	22.27%	0.161
One-off Validation	55.36%	0.00%	0.00%	0.00%	52.94%	82.35%	76.19%	84.21%	35.00%	48.94%	
One-Off Training	59.96%	0.00%	0.00%	19.40%	66.13%	87.36%	88.35%	56.35%	45.12%		

From Table 17 it can be seen that the linear regression model has zero accuracy in bins a1, a2 and a3. The scatter plot for this model is shown in Figure 29. Ideally, the best model would have an equal distribution of accuracy throughout the bin structure and be able to fit the full range of real delta VTR data, as well as predicting positive and negative values correctly.

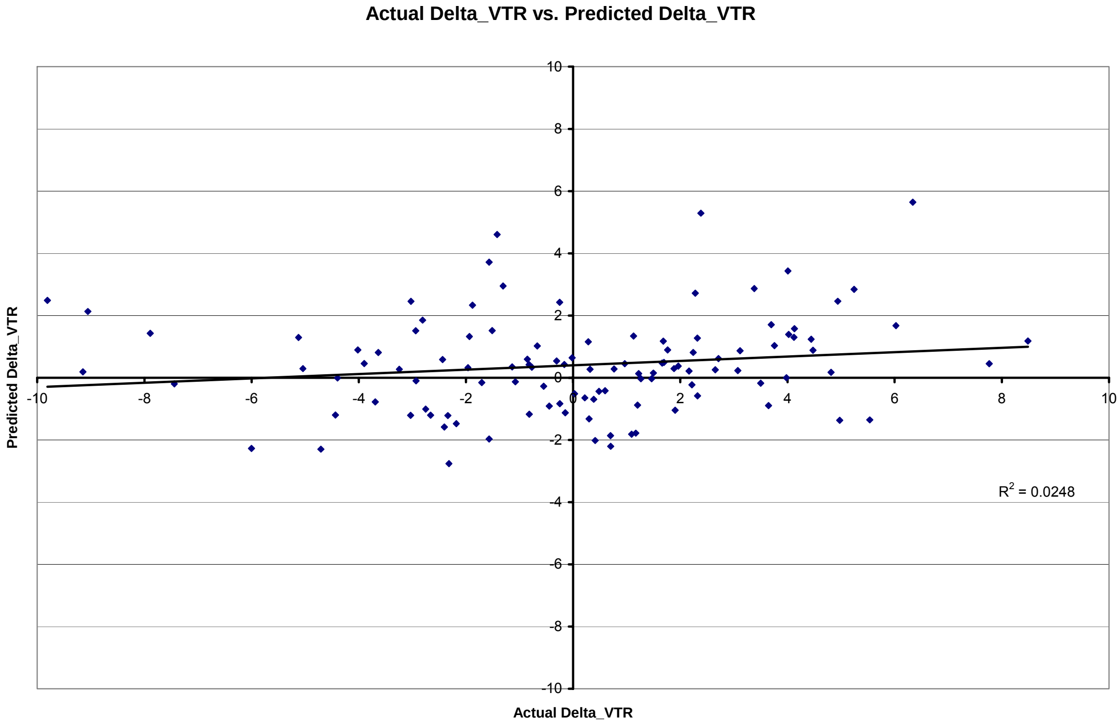


Figure 29: Scatter plot for forced enter linear regression model

The regression coefficients for this model are shown in Table 18.

Table 18: Variables and the associated forced enter regression coefficients

Variables	coefficient	t - value
(Constant)	-27.605	-1.749
Bus share	36.169	2.231
Carpool + Vanpool share	24.729	2.838
walk share	40.915	2.449
Motorcycle share	77.932	4.297
Bicycle share	96.179	4.393
Average Miles traveled	0.010	1.100
Average minutes traveled	-0.023	-0.287
3/36 compressed work week share	-0.004	-0.087
4/40 compressed work week share	0.147	0.661
8/80 compressed work week share	0.004	1.184
No. of employees	0.000	0.276
facilities & amenities(grouped incentives)	0.804	1.361
Guaranteed ride home programs(grouped incentives)	0.351	1.129
flexible timing (grouped incentives)	-0.484	-1.620
Ride share matching programs(grouped incentives)	-1.456	-0.342
financial incentives(grouped incentives)	-0.377	-0.680
Parking management (grouped incentives)	-0.692	-0.885
Telecommuting (grouped incentives)	0.465	1.453
Compressed work week(grouped incentives)	0.363	1.128
Vanpool vehicles(grouped incentives)	-0.673	-1.364
onsite incentives(grouped incentives)	0.116	0.339
Non financial (grouped incentives)	-0.028	-0.057
commuter tax benefit incentives(grouped incentives)	-0.385	-0.933
Vehicle trip rate	0.246	1.595

Phase II: Tucson Over-Sampled Data

To get better accuracies on the moderate range bins a2 to a5, the examples in these bins were over-sampled. Table 19 shows the changes in the number of examples in training set due to over-sampling.

Table 19: Tucson Over-Sampled Data – No. of Records in bins

Range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)							
		> - 7	[- 7 to - 4.5)	[- 4.5 to – 3)	[- 3 to -1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=
Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**
Validation	**112**	4	6	6	13	19	25	22	17
Training	**1612**	142	221	215	248	226	215	181	164

Neural network and regression models with different variable selection were built on this data. Figures 30, 31, and 32 show the comparison of the different performance measures on this data.

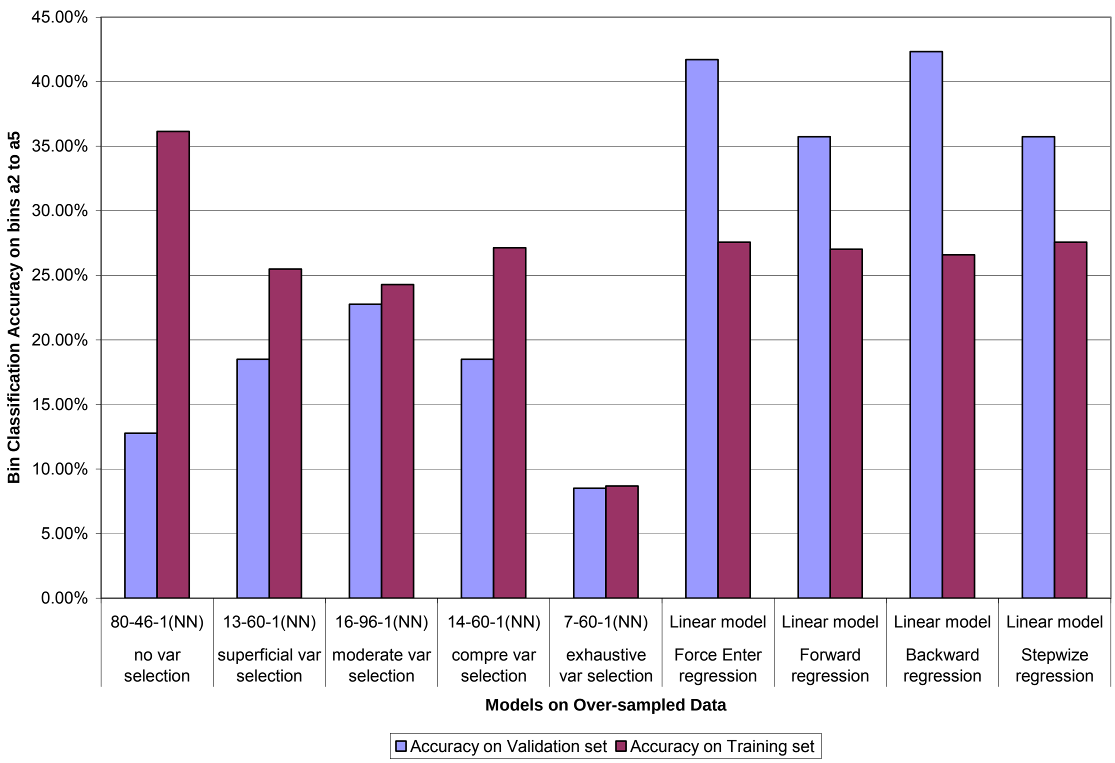


Figure 30: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on Over-sampled data)

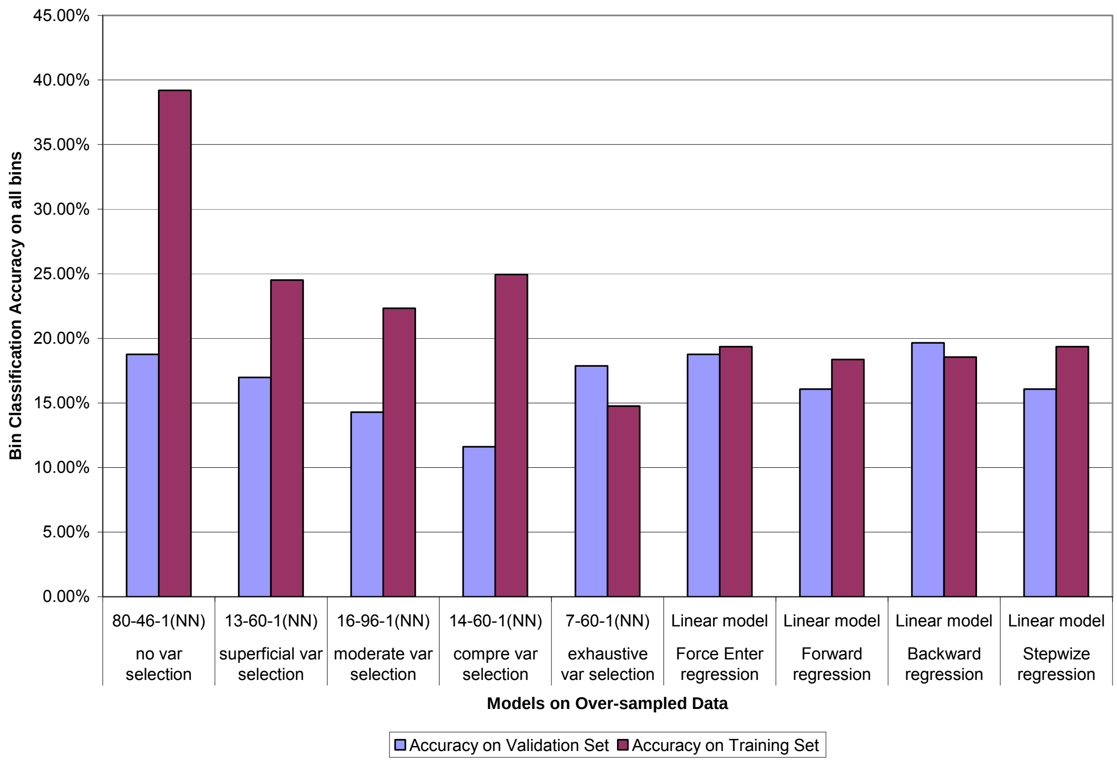


Figure 31: Bin Classification Accuracy on Full Range of change in VTR (all bins) for validation & training set (Different models on Over-sampled data)

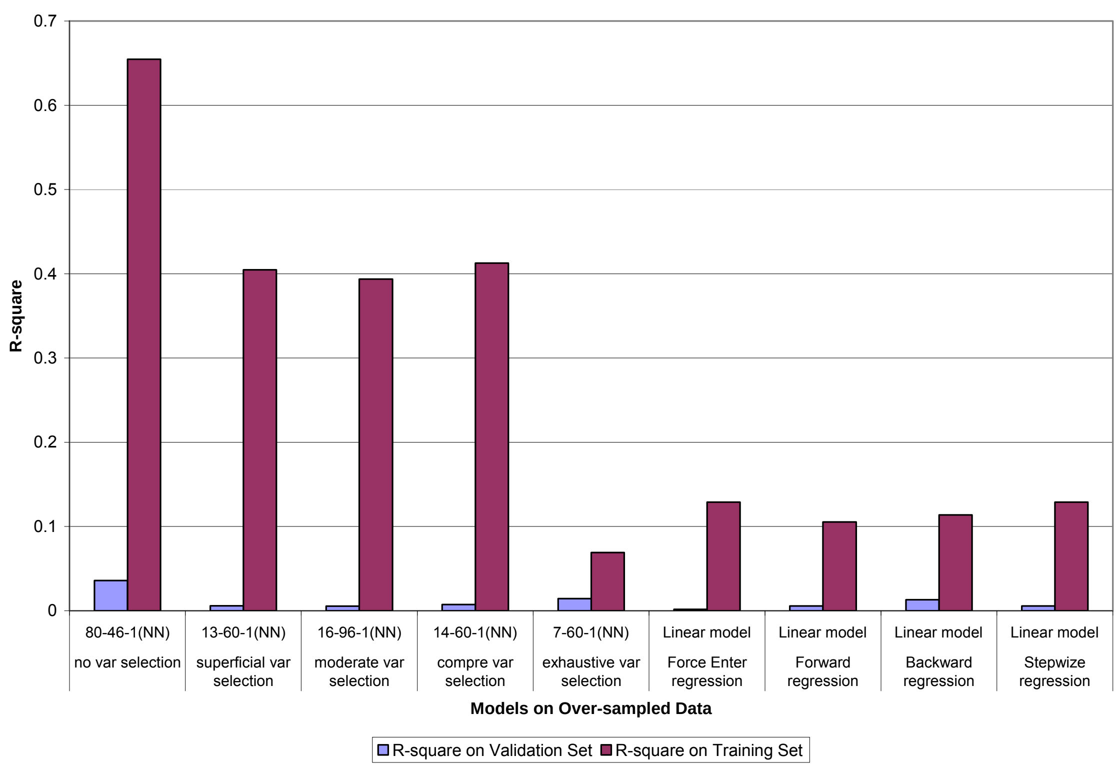


Figure 32: R-square for validation & training set (Different models on Over-sampled data)

The variables selected from the over-sampled datasets by these models are shown in Table 20

Table 20: Variables selected by different models on Tucson over-sampled data

variables	no var	Superficial	moderate	Comprehensive	Exhaustive	force enter	For ward	Back ward	Step wise
Alone share	√	√	√						√
Bus share	√		√	√	√	√	√	√	
Carpool + Vanpool share	√	√	√	√	√	√	√	√	
Walk share	√	√	√	√		√	√	√	
Motorcycle share	√	√	√	√		√	√	√	√
Bicycle share	√	√	√	√	√	√	√	√	√
Average Miles traveled	√			√	√	√		√	√
Average minutes traveled	√	√			√	√			
3/36 compressed work week share	√		√	√		√			
4/40 compressed work week share	√					√	√		√
8/80 compressed work week share	√		√			√		√	
Number of Employees	√		√			√		√	
Adjusted Work Hours									
Alternative Fuel Vehicles	√					√		√	√
Alternate Mode Campaign	√					√			√
Alternate Mode Information	√					√		√	
Bicycle Campaign	√					√			
Bicycle Racks	√					√			
Bus Pass Sales on Site						√			
Bus Subsidy						√			
Busing Vehicle	√					√			
Carpool Campaign	√					√		√	
Carpool Subsidy	√					√			
Carpooling Vehicle	√					√			
Compressed Work Week	√					√			
Coordination with Transit Provider	√	√	√	√		√		√	√
Covered Parking	√					√			
Daycare Facilities on Site	√					√			
Dissemination of Air Quality Information	√					√			
Drawing for Prizes	√	√				√			
Employee Shifts Between Sites	√					√	√		
Fee for Parking	√					√			
Field Worker	√					√			
Guaranteed Ride Home Program	√					√			
Incentive Programs	√					√			
Incentives for Employees to Live Close	√					√			
Information Center	√					√			
Map Board	√					√		√	√
Matching Service									
New Employee Information	√					√			
Newsletter Articles	√		√	√		√		√	√
Post Air Quality Information	√	√		√		√		√	
Preferred Parking									
Rebate not to Use Parking	√					√		√	
Rideshare Committee	√					√			
Showers/Lockers	√					√			
Shuttle Vehicle	√					√	√	√	
Speakers	√	√				√	√	√	√
Staging Area	√					√			
Transportation Fair						√			
Vanpooling Vehicle						√			
Vanpool Subsidy	√	√	√	√		√	√	√	√
Vanpooling Vehicle	√					√			
Walking Campaign	√				√	√			
Work at Home	√					√			
VTR	√		√	√		√		√	√

It can be seen from Table 20, that the variables selected by the neural network models with variable selection are not really the key incentives and therefore these models are deemed as unsuitable models. A neural network model built without variable selection (M2) was able to get the best R-square value with second best 'bin classification accuracy on full range of change in VTR'. But, it performed poorly in the required bins a2 to a5.

All regression models were able to get very good 'bin classification accuracy on moderate range of change in VTR' with backward regression model (M1) getting the best 'bin classification accuracy on full range of change in VTR' with second best R-square value. The detailed bins accuracies and scatter plots are shown in Table 21 and Figures 33 and 34.

Table 21: Detailed accuracies on bins: Tucson over-sampled data.

	range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
			> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to - 1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=		
	Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**		
	Validation	**112**	4	6	6	13	19	25	22	17	Weighted Avg on a2 to a5	R-square
	Training	**1009**	37	52	56	116	171	222	200	155		
M1	**Exact Validation**	19.64%	0.00%	25.00%	10.00%	47.06%	58.82%	9.52%	0.00%	0.00%	42.34%	0.013
	Exact Training	18.55%	0.00%	1.81%	26.98%	46.37%	28.76%	17.21%	7.73%	3.66%	26.59%	0.114
	One-off Validation	41.07%	0.00%	25.00%	50.00%	94.12%	82.35%	33.33%	15.79%	0.00%	75.53%	
	One-Off Training	48.82%	8.45%	25.79%	66.05%	91.13%	73.45%	55.81%	25.41%	10.98%		
M2	**Exact Validation**	18.75%	0.00%	0.00%	0.00%	17.65%	17.65%	19.05%	21.05%	35.00%	12.77%	0.036
	Exact Training	39.21%	67.61%	47.06%	34.88%	33.47%	29.65%	30.23%	33.70%	49.39%	36.15%	0.655
	One-off Validation	50.00%	20.00%	0.00%	50.00%	29.41%	64.71%	66.67%	52.63%	50.00%	43.62%	
	One-Off Training	78.35%	88.73%	87.33%	77.67%	79.03%	75.22%	73.95%	71.82%	74.39%		

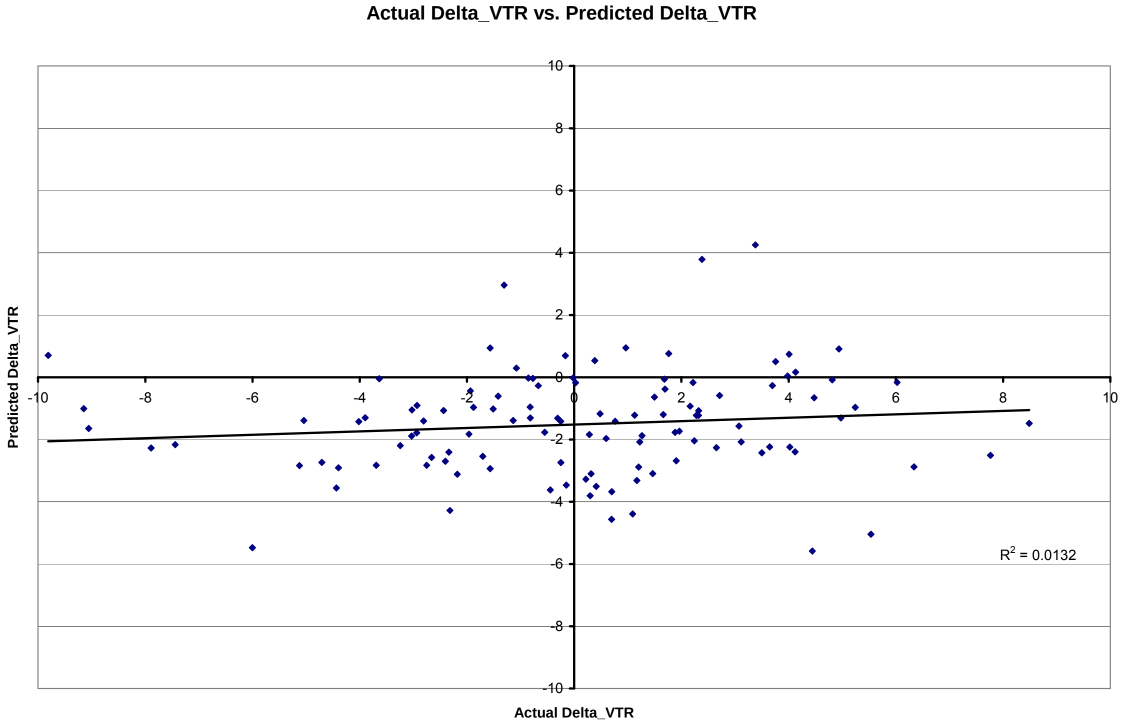


Figure 33: Scatter plot for linear backward regression model

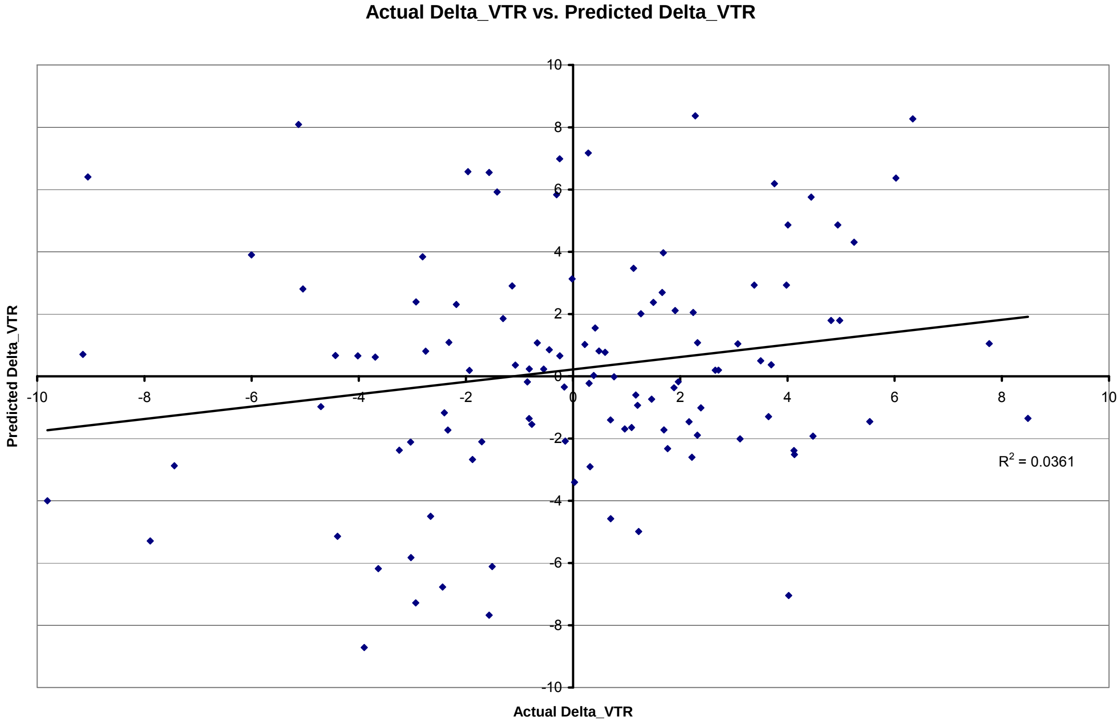


Figure 34: Scatter plot for neural network model built with no variable selection

It is apparent from the two scatter plots that the backward regression model has a very high tendency to predict negative changes in VTR, whereas the neural network model seems much more balanced, but still getting some predictions way off from the actual change in VTR. Also the neural network model gets better accuracy on the training set than the linear models and so in considered better than the linear model.

Table 22: Variable and backward regression coefficients

Variable	coefficients	t - value
(Constant)	-35.893	-2.923
Bus share	30.707	2.315
Carpool + Vanpool share	29.177	4.162
walk share	38.802	2.834
Motorcycle share	73.305	4.810
Bicycle share	97.354	5.230
Average Miles traveled	0.008	1.897
8/80 compressed work week share	0.007	1.790
No. of employees	0.000	1.913
Post Air Quality Information	0.711	2.085
Bus Pass Sales on Site	-0.476	-1.694
Alternative Fuel Vehicles	-1.099	-1.687
Compressed Work Week	0.797	2.977
Map Board	-0.696	-2.349
Preferred Parking	1.075	3.807
Rideshare Committee	0.964	3.424
Staging Area	-1.552	-1.936
Shuttle Vehicle	-1.494	-2.241
Showers/Lockers	0.763	2.851
Vanpooling Vehicle	-3.578	-2.983
Vehicle trip rate	0.292	2.352

Equivalent grouped incentives models were also built on this over-sampled data. Table 23 and Figures 35, 36, and 37 shows the accuracies and the R-square values of the neural network and the regression models built using no variable selection.

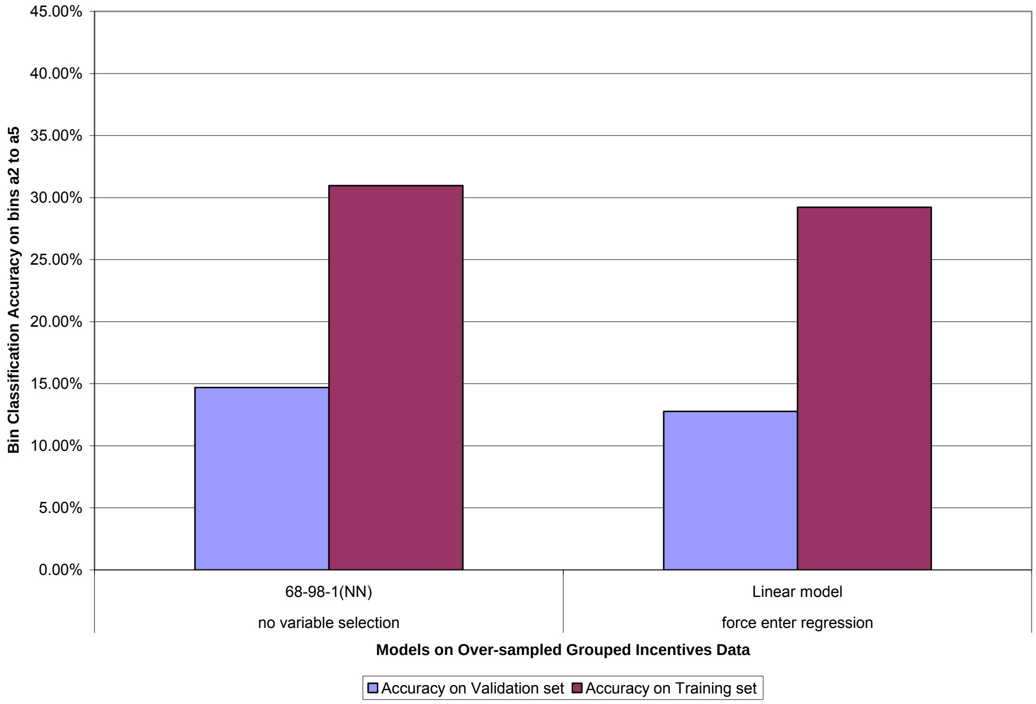


Figure 35: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on Over-sampled Grouped Incentive data)

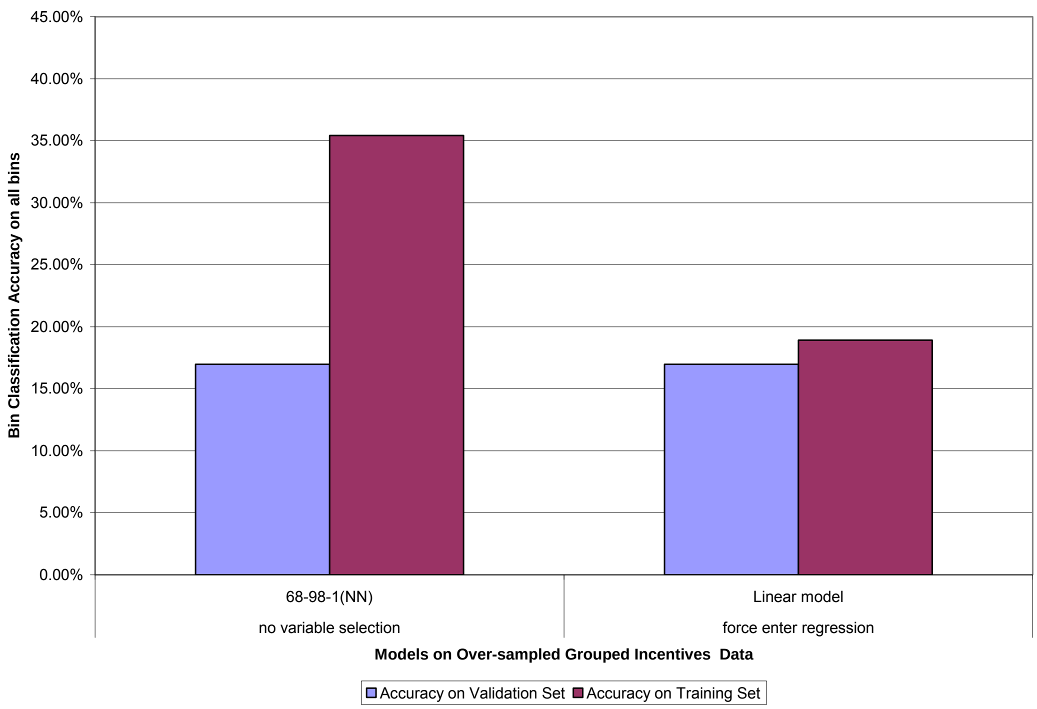


Figure 36: Bin Classification Accuracy on Full Range of change in VTR (all bins) for validation & training set (Different models on Over-sampled Grouped Incentive data)

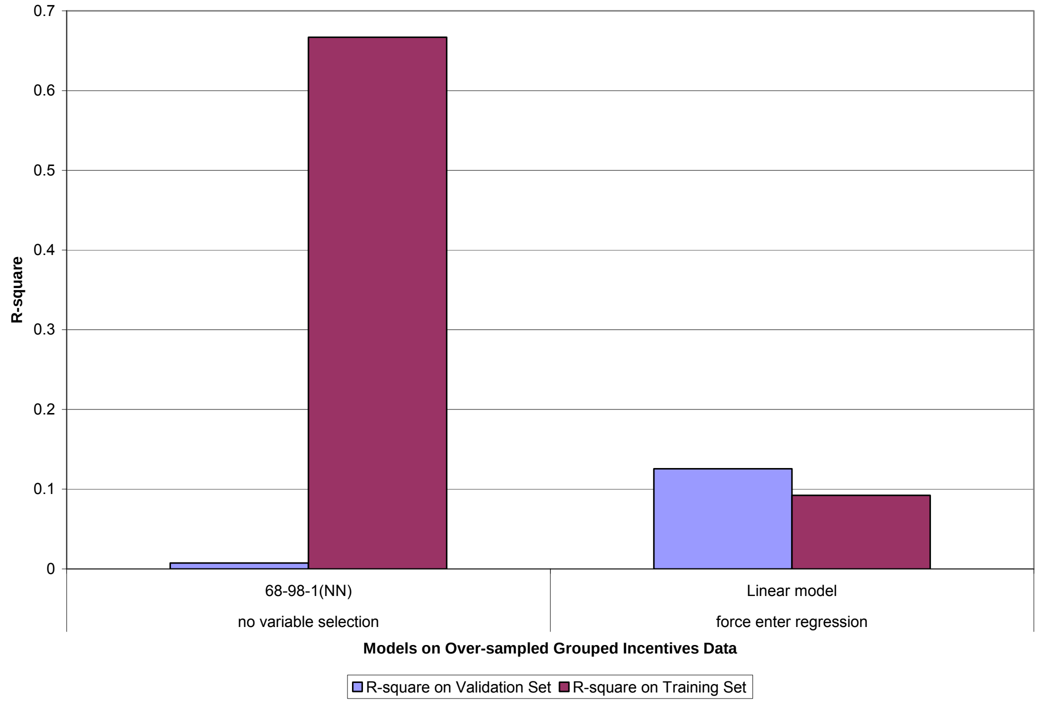


Figure 37: R-square for validation & training set (Different models on Over-sampled Grouped Incentive data)

For the validation set, it can be seen that both models have equivalent 'bin classification accuracy on full range of change in VTR' (16.96 percent), with the neural network model (M1) obtaining better accuracy in 'bin classification on moderate range of change in VTR' (14.68 percent) and the forced enter regression model (M2) having better R-square value (0.126) . The detailed bin accuracies are shown in Table 23 and the scatter plots are shown in Figures 38 and 39.

Table 23: Detailed accuracies on bins

	range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
			> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to - 1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=		
	Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**A6**	**a7**	**a8**		
	Validation	**112**	4	6	6	13	19	25	22	17	Weighted Avg on a2 to a5	R-square
	Training	**1009**	37	52	56	116	171	222	200	155		
M1	**Exact Validation**	16.96%	40.00%	0.00%	10.00%	17.65%	17.65%	19.05%	5.26%	25.00%	14.68%	0.007
	Exact Training	35.42%	67.61%	41.18%	29.30%	27.20%	26.67%	25.82%	30.22%	50.61%	30.96%	0.667
	One-off Validation	36.61%	40.00%	0.00%	30.00%	41.18%	41.18%	47.62%	31.58%	30.00%	35.53%	
	One-Off Training	76.24%	97.18%	71.04%	84.19%	73.60%	74.22%	72.77%	69.23%	73.78%		
M2	**Exact Validation**	16.96%	0.00%	0.00%	0.00%	5.88%	29.41%	42.86%	10.53%	10.00%	12.77%	0.126
	Exact Training	18.92%	0.00%	1.81%	26.98%	50.40%	34.96%	13.02%	3.31%	3.05%	29.23%	0.092
	One-off Validation	50.89%	0.00%	0.00%	0.00%	35.29%	100.00%	95.24%	52.63%	20.00%	48.94%	
	One-Off Training	47.77%	2.82%	12.67%	68.84%	95.16%	80.97%	54.88%	20.99%	9.15%		

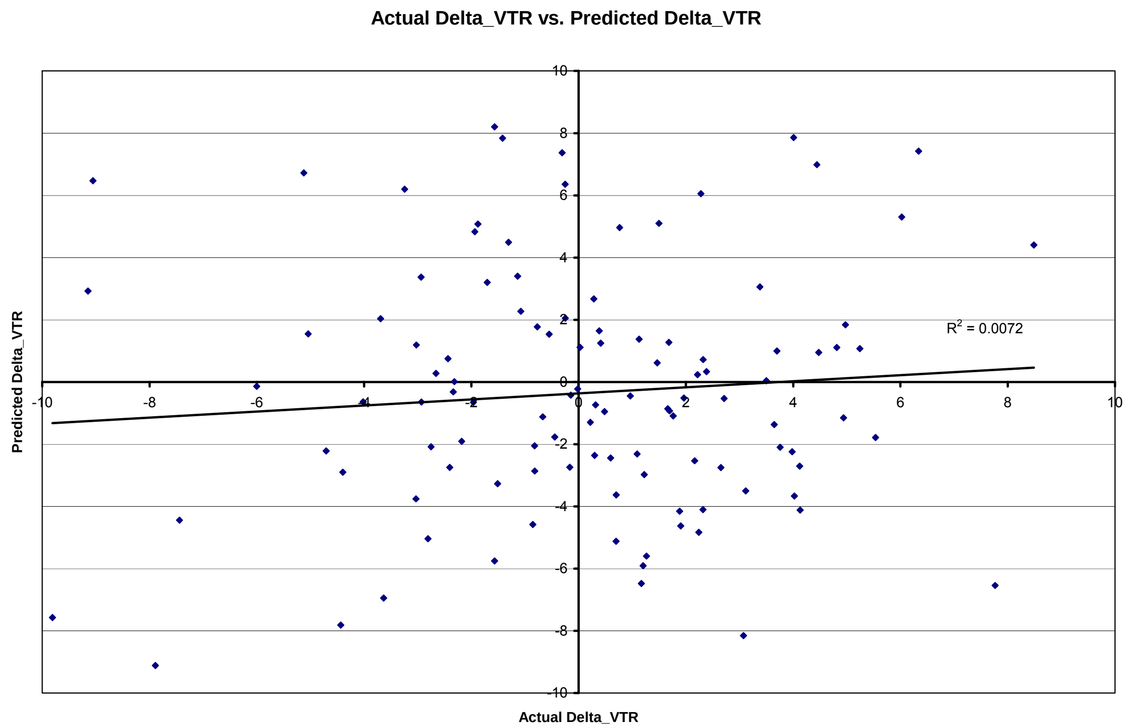


Figure 38: Scatter plot for neural network model

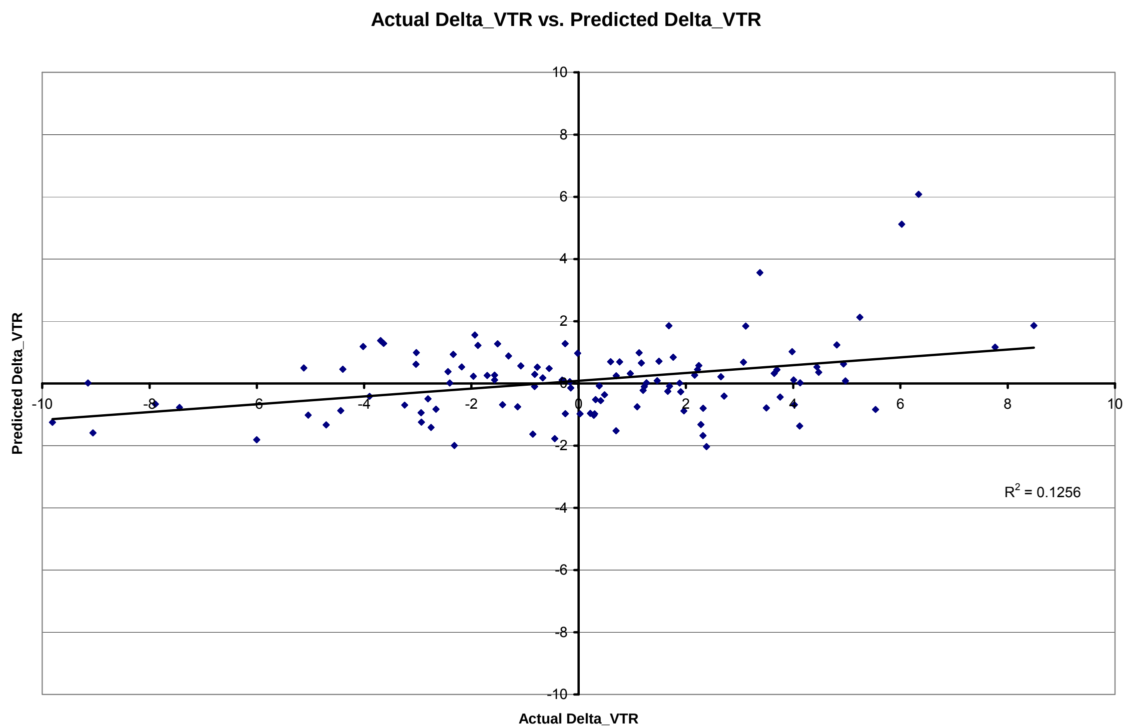


Figure 39: Scatter plot for linear forced enter regression model

The scatter plots in Figures 38 and 39 show that both the forced enter regression model and the neural net model with no variable selection are balanced in predicting the changes in VTR with regression model predicting smaller changes in VTR as opposed to neural network predicting large changes in VTR close to actual changes in VTR. The neural network model is able to get much better performance than linear regression model on training set and so can be considered better than the regression model.

The regression coefficients for the model are shown in Table 24.

Table 24: Variable and forced enter regression coefficients

Variables	coefficients	t - value
(Constant)	22.498	0.462
Bus share	-29.515	-0.577
Carpool + Vanpool share	3.275	0.121
walk share	-5.690	-0.105
Motorcycle share	-19.791	-0.352
Bicycle share	-52.474	-0.729
Average Miles traveled	-0.041	-1.622
Average minutes traveled	0.252	1.045
3/36 compressed work week share	-0.014	-0.067
4/40 compressed work week share	-2.557	-2.739
8/80 compressed work week share	0.007	1.213
No. of employees	0.002	2.614
facilities & amenities(grouped incentives)	-1.561	-0.910
Guaranteed ride home programs(grouped incentives)	-0.326	-0.407
flexible timing (grouped incentives)	-0.900	-1.232
financial incentives(grouped incentives)	-0.591	-0.343
Parking management (grouped incentives)	-3.646	-1.627
Telecommuting (grouped incentives)	-1.259	-1.550
Compressed work week(grouped incentives)	0.791	0.936
Vanpool vehicles(grouped incentives)	1.097	0.778
onsite incentives(grouped incentives)	1.425	1.630
Non financial (grouped incentives)	0.199	0.161
commuter tax benefit incentives(grouped incentives)	-1.735	-1.641
Vehicle trip rate	-0.235	-0.480

When neural network model built on grouped incentives with over-sampled data is compared with the neural network model built on ungrouped individual incentives with over-sampled data (Table 21 and 23), it can be seen that the model on grouped incentives has got accuracy distributed over all bins as against the other model for which the accuracy on bins a1, a2 and a3 is zero. Looking at other performance measures which are not much significantly different, it can be said that that grouped incentive model will be more preferred due to its simplicity

Recommended Model

None of the models were exceptionally better than the other models. All regression models were able to predict small changes in VTR as compared to the actual change in VTR. This shortcoming was not desirable. The two best candidate models were the neural network models without variable selection built, one with the ungrouped incentive variables on the full sample data and the other with the grouped incentives on the over-sampled data.

The reasons the neural network model built without variable selection on the full sample ungrouped incentive data was considered as candidate was (Table 21-M2)

1. It has the best 'bin classification accuracy on full range of change in VTR' of 20.54% which is much better than the random choice accuracy of 12.5%
2. It has 'bin classification accuracy on moderate range of change in VTR' of 16.6%
3. It has the 'R-square' value of 0.022
4. It includes all the variables in predicting change in VTR

The reasons the neural network model built without variable selection on the over-sampled grouped incentive data was considered as candidate was (Table 23-M1)

1. It is built on a simple grouped incentives variable set
2. It has 'bin classification accuracy on full range of change in VTR' of 16.96% which is better than the random choice accuracy of 12.5%
3. It has 'bin classification accuracy on moderate range of change in VTR' of 14.68%
4. It has a 'R-square' value of 0.007

Looking at the performance measures of both models, the neural network model built without variable selection on the full sample ungrouped incentive data can be said to be the recommended model. Also though is the model is not built on simple grouped incentives, it at least include all the incentives in predicting the changes in VTR.

WASHINGTON MODEL

Phase I: Washington Full Sample Data

The Washington dataset consisted of 1,414 total examples. Approximately ten percent of the dataset was left aside as validation set while the rest became training/testing set. For all of the Washington models the testing set was ten percent of the random data of the training/testing set. The eight bin ranges and the number of examples in each bin for validation and training/testing set are shown in Table 27.

Table 25: Washington Full Sample Data – No. of Records in bins

Range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)							
		> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to -1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=
Bin Number		**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**
Validation	**144**	15	15	15	17	20	20	21	21
Training	**1300**	136	132	132	152	182	184	190	192

Neural network models were built of data containing the entire variable set (105 variables) with different variable selections. This dataset contained variables such as shares of employee's incentive preference and type of work they performed. The employers might not be able to provide this information about their employees. So these variables with a few other insignificant variables were removed from the data. The models on the data containing these variables were built to study the contribution of these employee preference variables in predicting the change in VTR.

A neural network model with no variable selection and all different types of regression models were built on this condensed data. Figures 40, 41, and 42 show the different performance measures for all of these models.

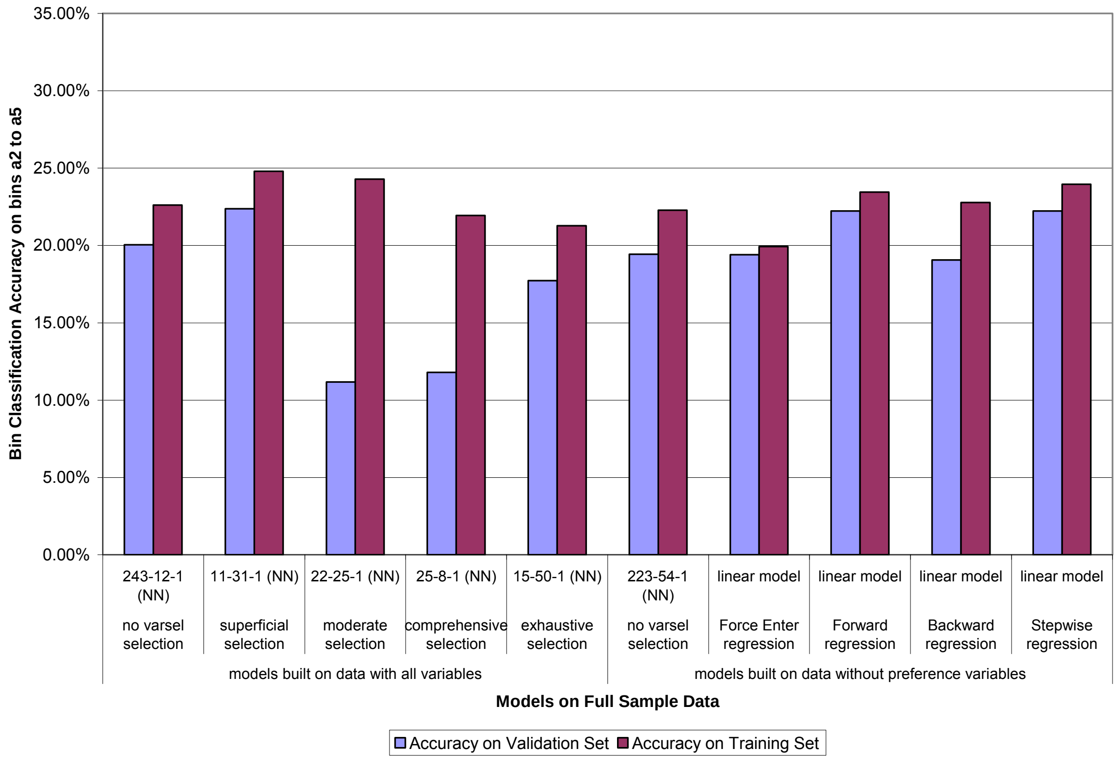


Figure 40: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on Full sample data)

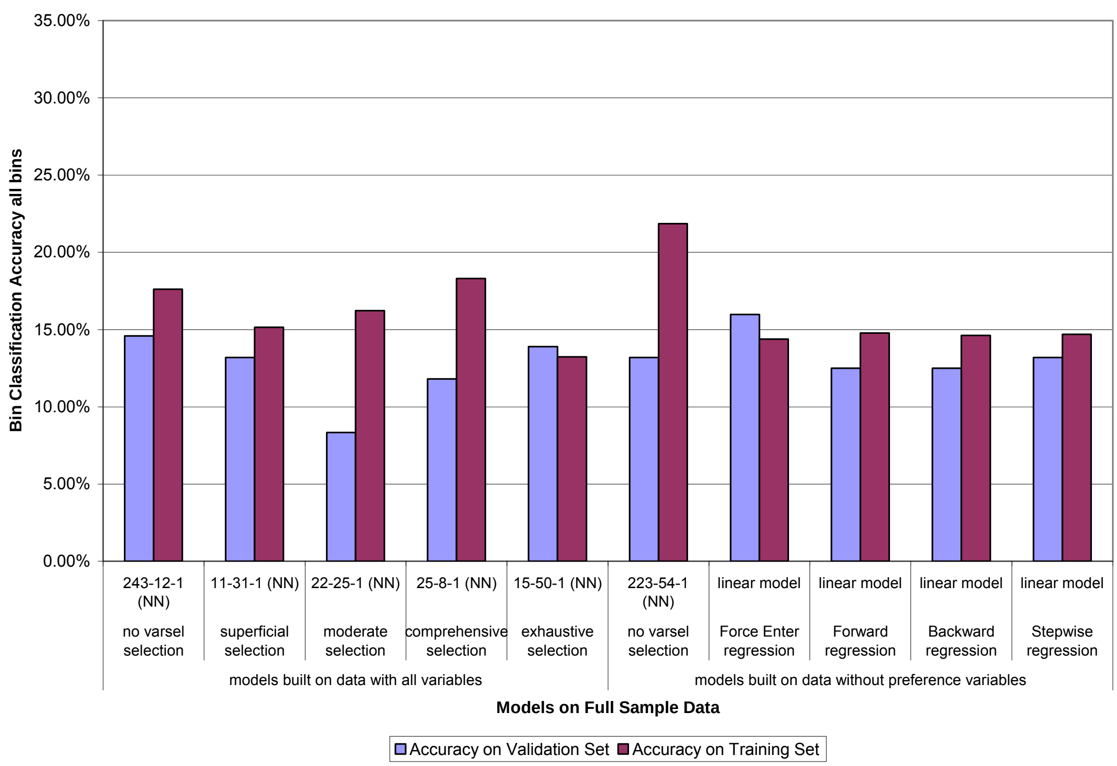


Figure 41: Bin Classification Accuracy on Full Range of change in VTR (all bins) for validation & training set (Different models on Full sample data)

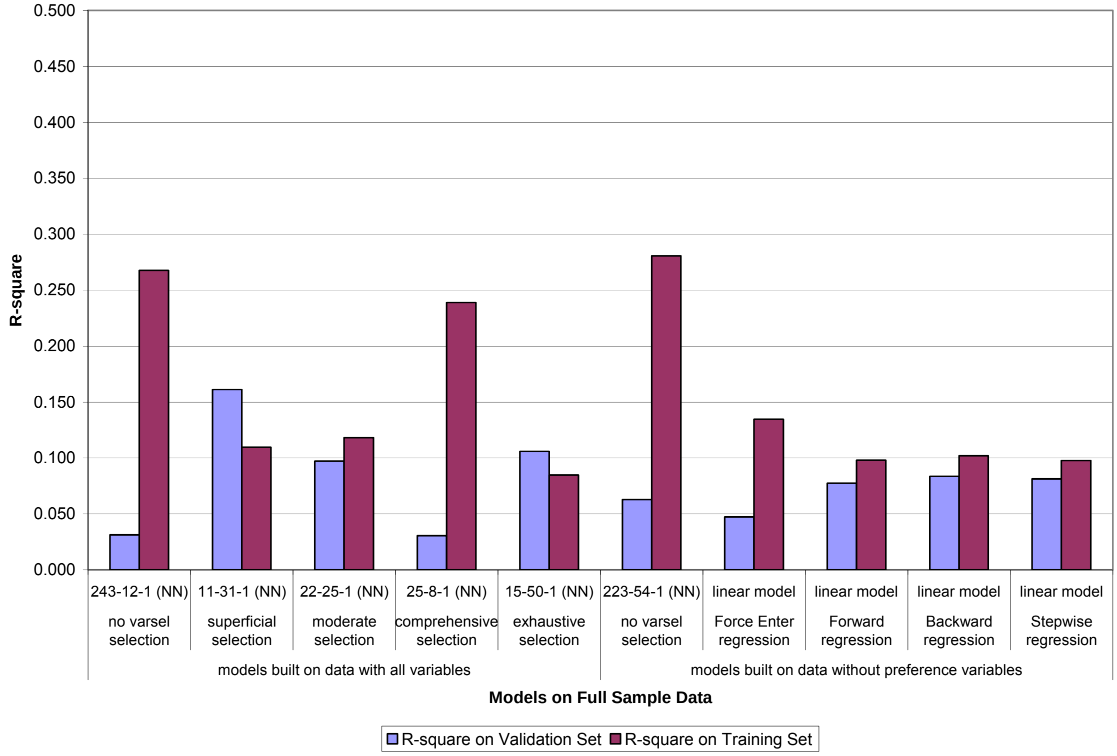


Figure 42: R-square for validation & training set (Different models on Full sample data)

Table 28 shows the variables selected by the different models.

Column name abbreviations,

N – Neural network model without variable selection
S – Neural network model with superficial variable selection
M – Neural network model with moderate variable selection
C – Neural network model with comprehensive variable selection
E – Neural network model with exhaustive variable selection
FE – Linear Forced Enter regression model
FR – Linear Forward regression model
BK – Linear Backward regression model
SP – Linear Stepwise regression model
√ - indicates the selection of the variable
'-' - indicates the variable was not present in the data

Table 26: Variables selected by different models on Full Sample data

Variables	NN models on data with all variables					Models on data without preference and other insignificant variables				
	N	S	M	C	E	N	FE	FR	BK	SP
Non-profit organization	√					-	-	-	-	-
Agriculture organization						-	-	-	-	-
Finance organization	√					-	-	-	-	-
Info services organization	√					-	-	-	-	-
Health organization	√					√	√			
Retail organization	√					√	√			
Manufacturing organization	√					√	√		√	
Services organization	√					-	-	-	-	-
Public utilities organization	√		√			-	-	-	-	-
Construction organization						-	-	-	-	-
Transportation organization	√					-	-	-	-	-
Government organization	√		√			-	-	-	-	-
Other organization	√					√	√		√	
Offered to All	√					-	-	-	-	-
Union	√					√	√			
Shifts	√					√	√			
Onsite Parking Spaces	√					√	√			
Offsite Parking Spaces	√	√		√		√	√			
Leased Onsite Parking Price	√				√	√	√			
Leased Offsite Parking Price	√		√			-	-	-	-	-
Own Onsite Parking	√		√	√		√	√	√	√	√
Onsite Parking Charge	√					√	√			
Own Offsite Parking	√					√	√			
Offsite Parking Charge	√			√		√	√			
Pay Parking Charge	√					√	√			
On off parking sub	√			√		√	√	√	√	√
Free Parking 1/4 mile	√					√	√			
ETC Additional Training	√					√	√			
ETC Onsite	√					√	√			
Distribute Info	√					√	√			
Post Materials	√				√	√	√			
CTR Orientation	√					√	√			
CTR Events	√			√	√	√	√	√	√	√
CTR E-mail	√			√		√	√			
Articles	√					√	√			
Ride match Apps	√			√		√	√			
With Paychecks	√				√	√	√			
Drawings	√	√		√		√	√	√	√	√
Leave						-	-	-	-	-
Other Promo	√					√	√			
Covered Bike Number	√					√	√			
Uncovered Bike Number	√					√	√			
Lockers Number	√					√	√			
Showers Number	√			√		√	√			
Shelters Number	√					√	√			
Other Amenities 1 Number	√		√			√	√	√	√	√
Carpool Spaces Number	√			√		√	√			
Vanpool Spaces Number	√					√	√			
SOV Parking Charge	√					√	√		√	
SOV Parking Charge Number	√			√		√	√			
Reduced SOVP Number	√					√	√			
Transit Subsidy	√		√		√	√	√			
Ferry Subsidy	√				√	√	√			
Vanpool Subsidy	√					√	√			
Carpool Subsidy	√					√	√			
Walking Subsidy	√					√	√			
Bike Subsidy	√					√	√			
Employees on flextime	√					√	√			
Employees with GRH	√					√	√			
Employees in-house match	√					√	√			
Employees public match	√					√	√			

Variables	NN models on data with all variables					Models on data without preference and other insignificant variables				
	N	S	M	C	E	N	FE	FR	BK	SP
FV work employees	√					√	√			
Number of Employees	√			√		√	√			
Percentages of employees on 3/36 CWW	√				√	√	√			
Percentages of employees on 4/40 CWW	√	√				√	√	√		
Percentages of employees on 5/40	√					√	√			
Percentages of employees on 7/40 CWW	√			√		√	√			
Percentages of employees on 9/80 CWW	√	√				√	√			
Percentages of employees on other CWW	√		√	√		√	√	√	√	√
Alone share	√			√	√	√		√		√
Bike share	√		√	√		√	√	√		√
Bus share	√	√			√	√	√		√	
Cars hare	√		√	√		√	√	√		√
Other share	√	√	√			√	√			
Tele share	√					√	√		√	
Van share	√					√	√	√		√
Walk share	√	√			√	√	√		√	
Days saved telecommuting in two weeks	√					-	-	-	-	-
Administration job Share	√			√		-	-	-	-	-
Craft/Production/Labor Share	√		√			-	-	-	-	-
Management job Share	√					-	-	-	-	-
Sales/Marketing job Share	√					-	-	-	-	-
Customer Service job Share	√		√			-	-	-	-	-
Other job Share	√		√			-	-	-	-	-
Professional/Technical job Share	√		√			-	-	-	-	-
Q8_Invalid_Share	√				√	-	-	-	-	-
Prefer provide car for work	√					-	-	-	-	-
Employee Prefer Transport during lunch	√		√	√		-	-	-	-	-
Employee Prefer GRH	√					-	-	-	-	-
Employee Prefer flex to meet CVpool bus	√					-	-	-	-	-
Employee Prefer financial incentive	√					-	-	-	-	-
Employee Prefer reserved discounted CVpool space	√	√		√		-	-	-	-	-
Employee Prefer Personalized help for CVpool	√		√	√	√	-	-	-	-	-
Employee Prefer covered bicycle parking	√					-	-	-	-	-
Employee Prefer lockers & showers	√					-	-	-	-	-
Employee Prefer onsite childcare	√		√			-	-	-	-	-
Employee Prefer CWW	√				√	-	-	-	-	-
Employee Prefer to telecommute	√					-	-	-	-	-
Employee Prefer improved access to transit	√		√		√	-	-	-	-	-
SOV	√	√	√	√	√	-	-	-	-	-
VMT	√	√		√		-	-	-	-	-
VTR	√		√			√	√	√	√	√
Central Business District	√					√	√			
Suburban area	√					√	√			
Outside suburban area	√					√	√			

From the plots in Figures 40 and 41, it is apparent that the model built using superficial variable selection obtained the best ‘bin classification accuracy in moderate range of change in VTR’ (22.37 percent), the best R-square value (0.16) and the ‘best accuracy on full range of change in VTR. But from Table 28 it can be seen that this model selected very few insignificant incentives that were considered as unsuitable. All neural net models built with all of the variables, selected many of the employee’s preferences for incentives. This shows that the preferences did play an important role in predicting the change in VTR. As stated earlier, the

employers might not be able to have access to this employee information; and so these models will be used simply for study.
Considering all of the performance measures and the variables included, the linear forced enter regression model (M1) and neural network model (M2) built on data without preference variables can be considered suitable. The accuracies for these models on each of the bins are shown in Table 29.

Table 27: Detailed accuracies on bins

	Range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
			> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to - 1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=		
	Bin Number		**a1**	**a2**	**a3**	**a4**	**A5**	**a6**	**A7**	**a8**		
	Validation	**144**	15	15	15	17	20	20	21	21	Weighted Avg on a2 to a5	R-square
	Training	**1300**	136	132	132	152	182	184	190	192		
M1	**Exact Validation**	15.97%	5.88%	11.76%	8.33%	25.00%	31.25%	23.08%	5.56%	13.64%	19.40%	0.047
	Exact Training	14.38%	2.84%	5.00%	11.27%	34.00%	27.27%	18.89%	10.00%	6.13%	19.93%	0.135
	One-off Validation	38.89%	17.65%	23.53%	25.00%	75.00%	75.00%	53.85%	11.11%	27.27%	50.00%	
	One-Off Training	46.08%	9.93%	28.57%	52.82%	80.67%	80.61%	59.44%	34.12%	24.06%	19.40%	0.047
M2	**Exact Validation**	13.19%	0.00%	11.76%	16.67%	31.25%	18.75%	15.38%	5.56%	9.09%	19.43%	0.063
	Exact Training	21.92%	6.38%	14.29%	18.31%	32.00%	27.27%	25.00%	17.06%	29.72%	23.28%	0.281
	One-off Validation	38.89%	17.65%	35.29%	33.33%	62.50%	43.75%	61.54%	27.78%	22.73%	43.78%	
	One-Off Training	53.38%	26.95%	37.14%	50.70%	66.00%	63.64%	66.67%	48.24%	59.43%		

The scatter plots for the validation set for these models are shown in Figures 43 and 44.

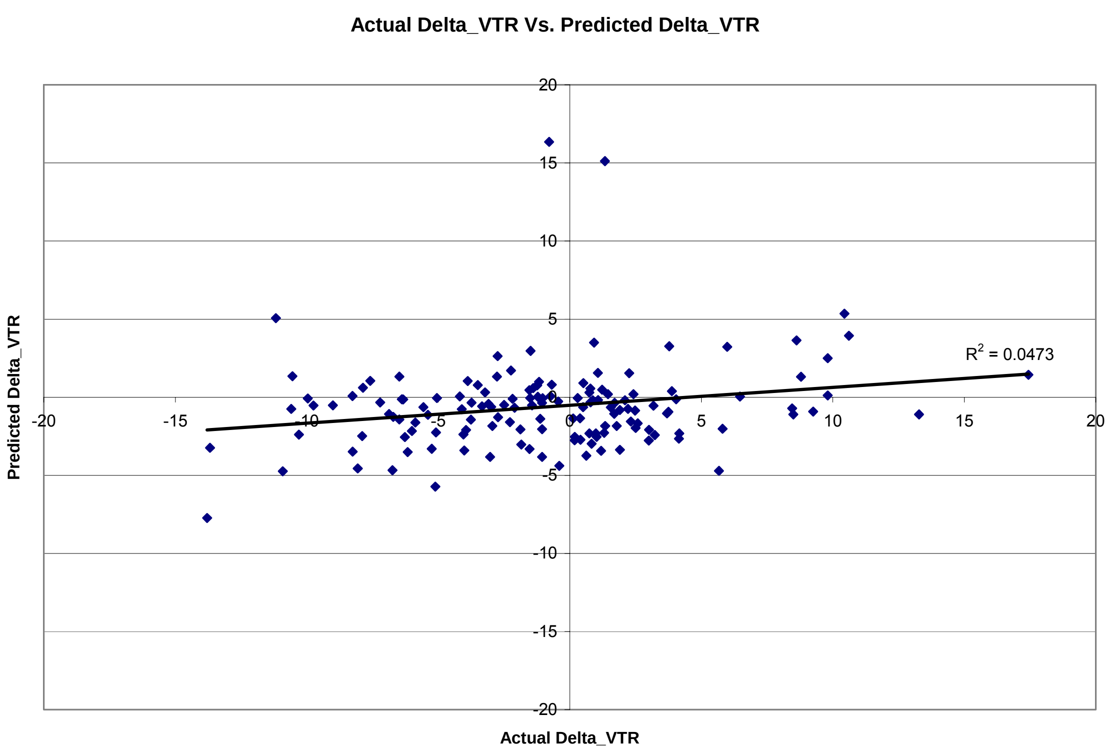


Figure 43: Scatter plot for linear forced enter regression model (M1)

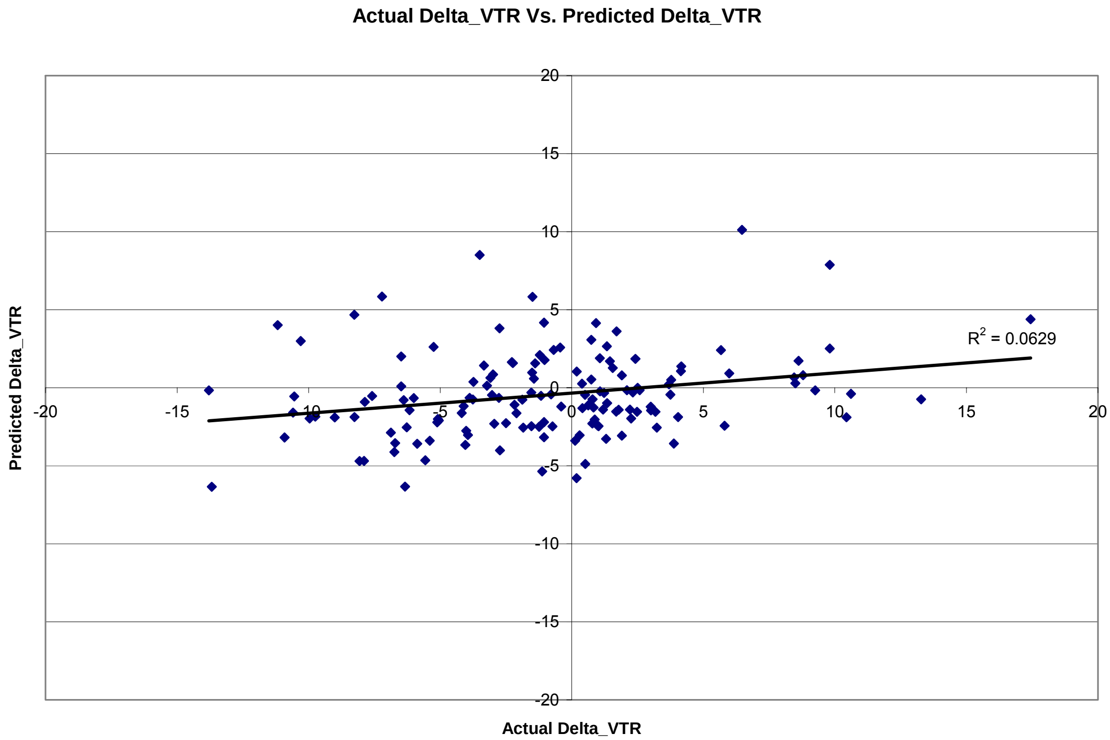


Figure 44: Scatter plot for neural network model (M2)

The regression coefficients for the forced enter regression model (M1) are shown in Table 30.

Table 28: Variables and coefficients for forced enter regression model

Variable	coefficient	t - value
Constant	49.063	2.267
Health organization	-0.386	-0.419
Retail organization	-1.695	-1.222
Manufacturing organization	1.337	1.637
Other organization	0.514	0.965
Shifts	0.547	1.463
Onsite Parking Spaces	-0.129	-0.672
Offsite Parking Spaces	-0.993	-1.794
Leased Onsite Parking Price	-0.002	-0.508
Leased Offsite Parking Price	-0.006	-1.130
Own Onsite Parking	0.956	2.412
Onsite Parking Charge	-0.004	-0.327
Own Offsite Parking	-0.254	-0.395
Offsite Parking Charge	0.010	1.103
Pay Parking Charge	0.007	1.275
on off parking sub	-1.793	-2.587
Free Parking 1/4 mile	0.321	1.142
ETC Additional Training	0.170	0.507
ETC Onsite	0.040	0.088
Distribute Info	0.005	1.253
Post Materials	0.001	0.121
CTR Orientation	0.000	0.665
CTR Events	-0.125	-2.366
CTR E-mail	-0.005	-0.390
Articles	0.017	0.486
Ridematch Apps	-0.002	-0.678

Variable	coefficient	t - value
With Paychecks	0.038	0.899
Drawings	0.071	2.147
Other Promo	0.004	0.664
Covered Bike Number	0.183	0.155
Uncovered Bike Number	3.276	1.024
Lockers Number	-0.018	-0.077
Showers Number	-7.013	-1.348
Shelters Number	-2.191	-0.411
Other Amenities 1 Number	-24.724	-3.366
Carpool Spaces Number	-1.593	-1.556
Vanpool Spaces Number	4.003	1.121
SOV Parking Charge	-0.012	-1.935
SOV Parking Charge Number	0.693	1.698
Reduced SOVP Number	2.695	1.217
Transit Subsidy	-0.002	-0.349
Ferry Subsidy	0.021	1.575
Vanpool Subsidy	-0.024	-1.770
Carpool Subsidy	0.007	0.555
Walking Subsidy	0.023	0.673
Bike Subsidy	-0.006	-0.322
employees on flextime	0.518	0.700
employees with GRH	0.241	0.655
employees in-house match	-0.200	-1.209
employees public match	-0.126	-1.257
FV work employees	-0.013	-0.006
Number of Employee	0.000	-0.512
Percentages of employees on 3/36 CWW	-0.621	-0.056
Percentages of employees on 4/40 CWW	-4.976	-1.085
Percentages of employees on 5/40	-1.665	-0.675
Percentages of employees on 7/40 CWW	-6.289	-0.843
Percentages of employees on 9/80 CWW	-4.998	-1.562
Percentages of employees on other CWW	-18.379	-2.609
Bike share	-18.862	-0.856
Bus share	-45.089	-2.274
Cars hare	-14.707	-1.280
Other share	-30.923	-1.372
Tele share	-47.565	-1.744
Van share	-20.061	-0.844
Walk share	-50.286	-2.490
VTR	-0.499	-2.470
Central Business District	-3.150	-0.577
Suburban area	-2.601	-0.525
Outside suburban area	-2.070	-0.449

To reduce to complexity of the models, all of the individual incentives were replaced with the grouped incentives and simple neural network and regression model with no variable selection were built on the data just containing the grouped incentives and worksite characteristics like mode-splits. The variables for the grouped incentives and worksite characteristics are shown in Table 31.

Table 29: Variables with grouped incentives

Variables	Description
Health	Health organization
Retail	Retail organization
Manufact	Manufacturing organization
other new	other organization
Shifts	Shifts
FACILITY_AMENITIES	Covered/uncovered bike parking, lockers a& showers, shelters, carpool & vanpool spaces, other amenities
GRH	Guaranteed ride home program
FLEX	flexible timing
Mrkt	CTR Events, CTR E-mail, Articles, With Paychecks, Other Promo, etc
RS_MATCH	Ride match Apps, employees in-house match, employees public match
FINANCIAL	Ferry, Carpool, Walking, Bike Subsidy
PARKMGT	SOV Parking Charge, SOV Parking Charge Number, Reduced SOVP Number
CWW	compressed work week 3/36, 4/40, 9/80, 7/40, other
onsite	onsite incentives
direct_nonfinan	Drawing, leaves, etc
commtax	Transit, vanpool subsidy
EmpNo	Number of employees
CWW3	Percentages of employees on 3/36 CWW
CWW4	Percentages of employees on 4/40 CWW
CWW5	Percentages of employees on 5/40
CWW7	Percentages of employees on 7/40 CWW
CWW9	Percentages of employees on 9/80 CWW
CWWOTHER	Percentages of employees on other CWW
AloneShr	Alone share
BikeShr	Bike share
BusShr	Bus share
CarShr	Cars hare
OtherShr	Other share
TeleShr	Tele share
VanShr	Van share
WalkShr	Walk share
VTR	VTR
CBD	Central Business District
Suburb	Suburban area
Outside	Outside suburban area

Figures 45, 46, and 47 show the comparison of the different performance measures on this grouped incentive data.

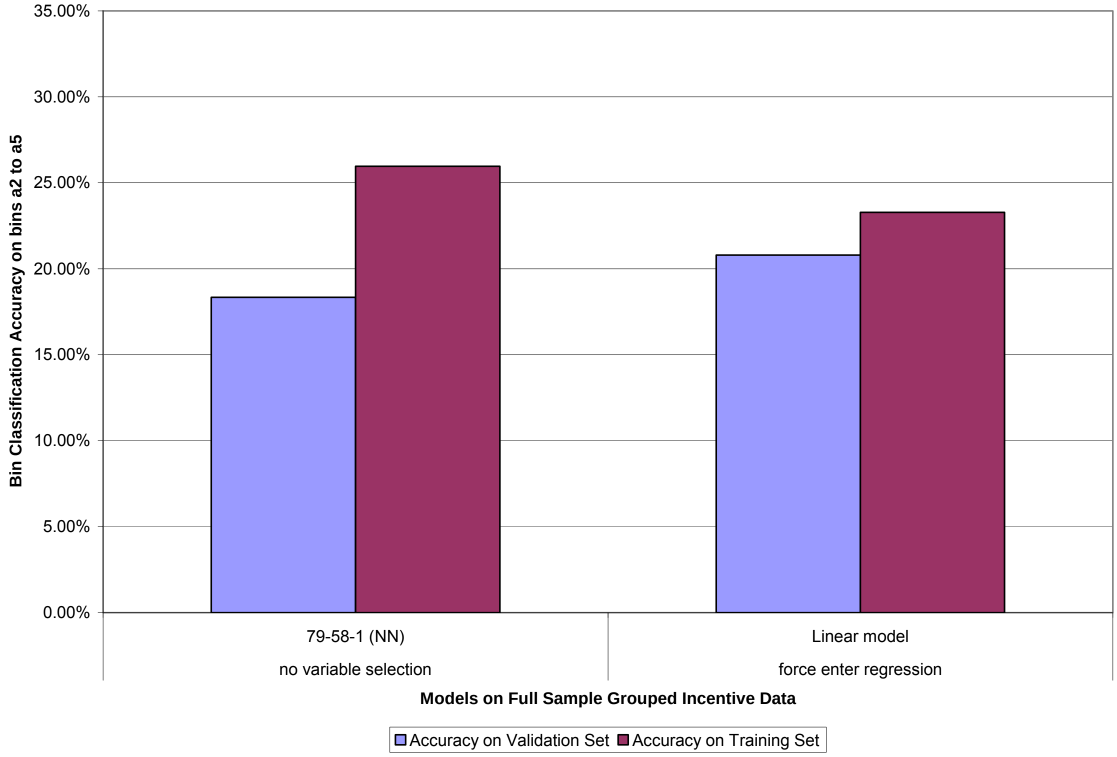


Figure 45: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on Full sample Grouped Incentive data)

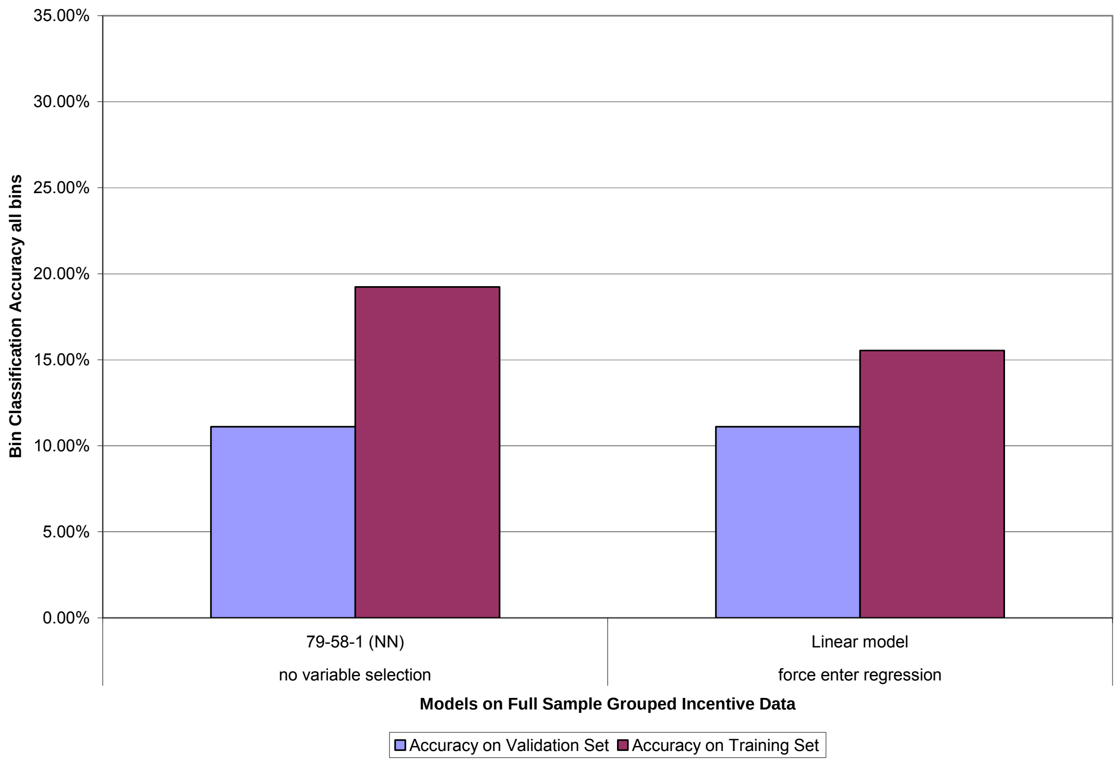


Figure 46: Bin Classification Accuracy on Full Range of change in VTR (all bins) for validation & training set (Different models on Full sample Grouped Incentive data)

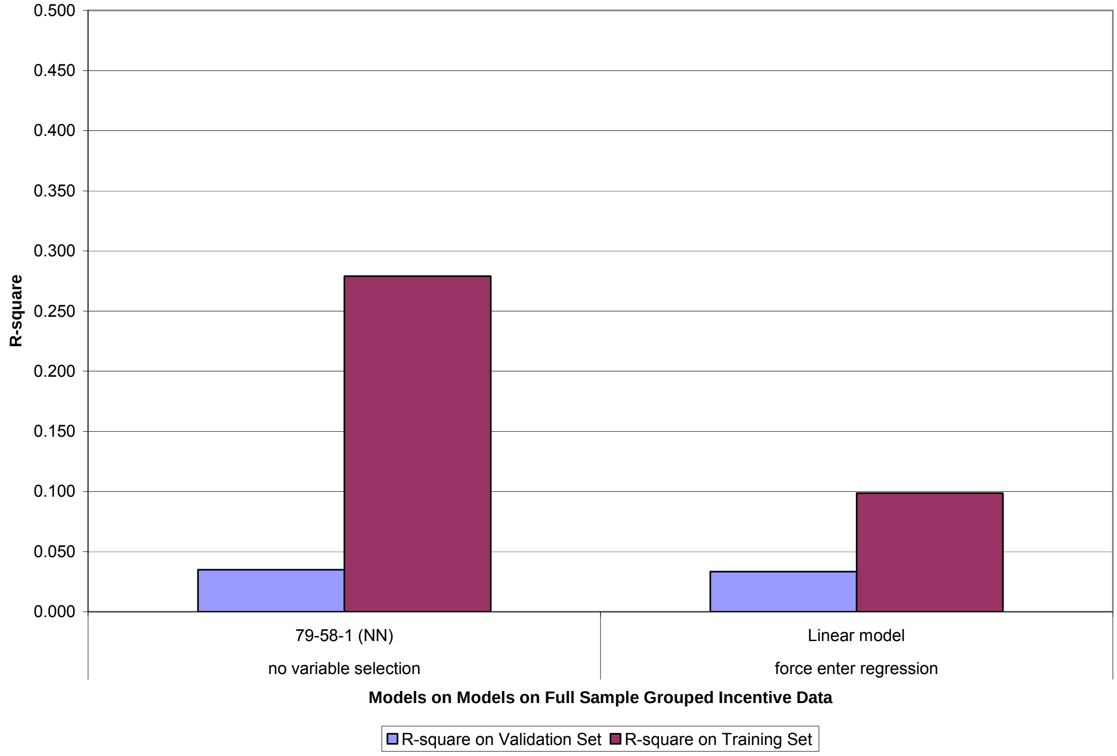


Figure 47: R-square for validation & training set (Different models on Full sample Grouped Incentive data)

The charts in Figures 45, 46, and 47 show that both of the models, neural network (M1) and linear forced enter regression (M2), are comparable to each other for 'bin classification accuracy on full range of change in VTR' but reduced as compared to models built on earlier ungrouped data. Even though the new linear forced enter regression model that was built on grouped incentives was better distributed than the previous model built with ungrouped incentives, the linear forced enter regression (M2) model was still not as well distributed between the bins as the neural network (M1) model. The detailed accuracies for both the models are shown in Table 32.

Table 30: Detailed accuracies on bins

	range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
			> - 7	[- 7 to – 4.5)	[- 4.5 to - 3)	[- 3 to - 1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=		
	Bin Number		a1	A2	a3	a4	A5	a6	a7	a8		
	Validation	144	15	15	15	17	20	20	21	21	Weighted Avg on a2 to a5	R-square
	Training	1300	136	132	132	152	182	184	190	192		
M1	Exact Validation	11.11%	0.00%	29.41%	25.00%	6.25%	12.50%	11.54%	0.00%	9.09%	18.34%	0.035
	Exact Training	19.23%	16.31%	26.43%	28.87%	25.33%	23.64%	13.33%	8.82%	15.57%	25.96%	0.279
	One-off Validation	36.81%	17.65%	52.94%	66.67%	68.75%	37.50%	30.77%	11.11%	27.27%	55.70%	
	One-Off Training	48.69%	46.81%	53.57%	76.06%	61.33%	62.42%	38.33%	28.24%	33.96%	18.34%	
M2	Exact Validation	11.11%	0.00%	0.00%	0.00%	31.25%	50.00%	3.85%	0.00%	9.09%	20.80%	0.033
	Exact Training	15.54%	0.00%	0.71%	9.86%	39.33%	39.39%	19.44%	8.24%	6.60%	23.28%	0.099
	One-off Validation	34.72%	0.00%	5.88%	50.00%	62.50%	75.00%	53.85%	11.11%	22.73%	47.34%	
	One-Off Training	42.23%	2.13%	22.14%	49.30%	84.67%	77.58%	58.33%	26.47%	18.87%		

The scatter plots for the validation set of the forced enter regression model and the neural net model with no variable selection are shown in Figures 48 and 49.

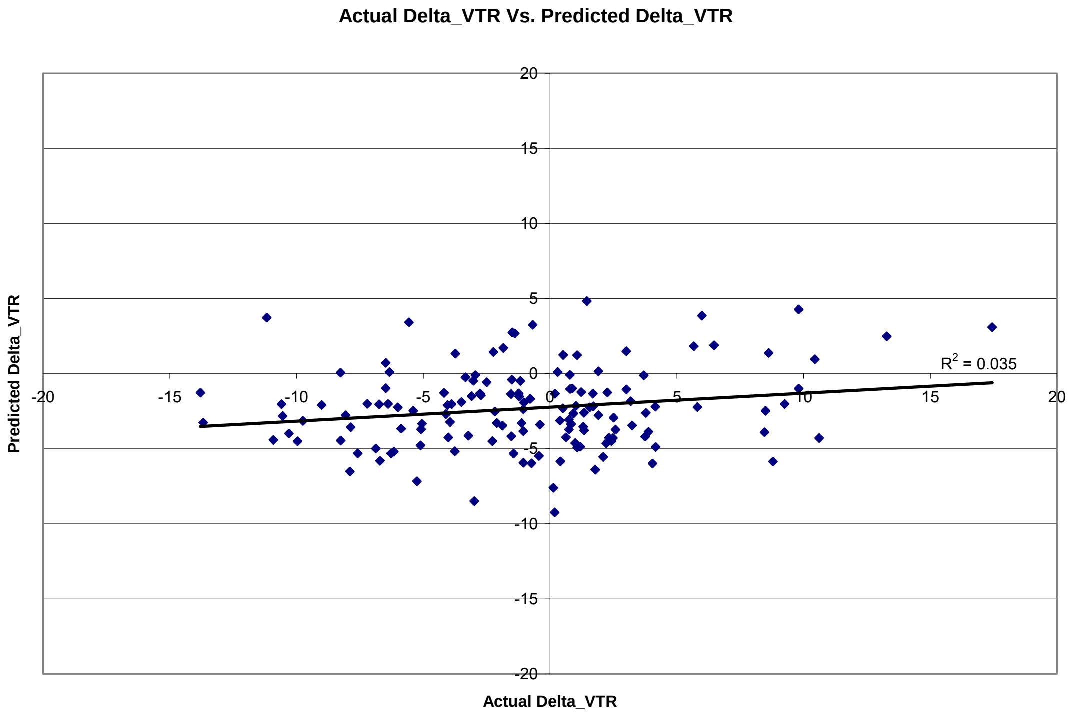


Figure 48: Scatter plot for neural network model (M1)

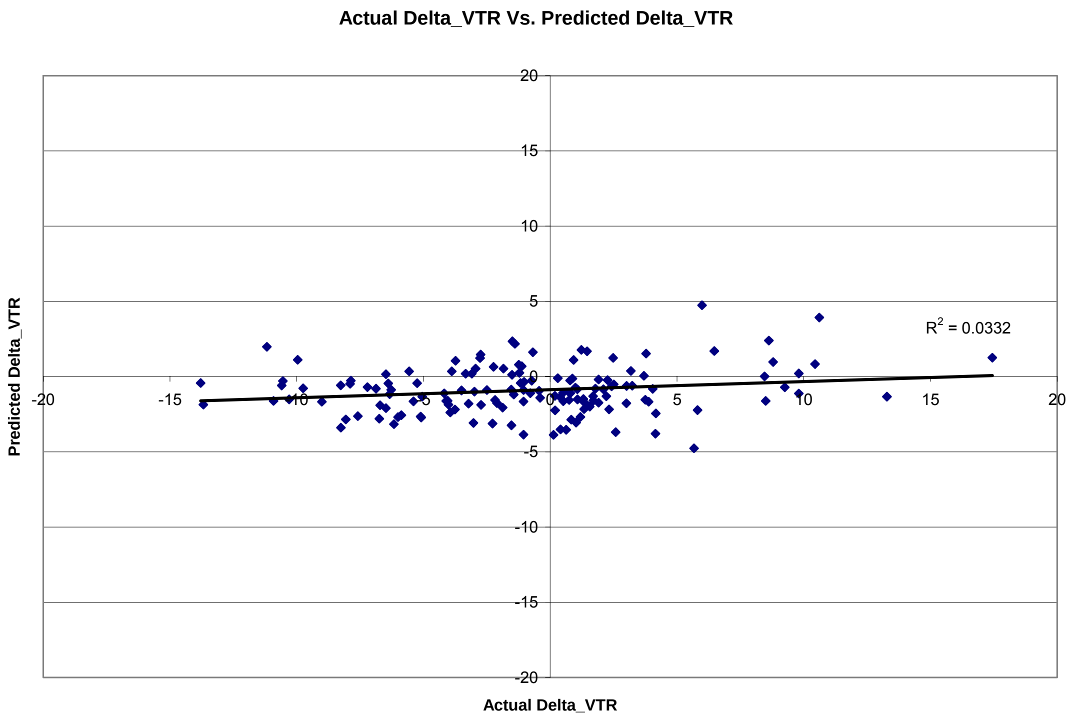


Figure 49: Scatter plot for linear forced enter regression model (M2)

The regression coefficients for the linear forced enter regression model (M2) (grouped incentives) are shown in table 33.

Table 31: Variables and coefficients for forced enter regression model

Variables	coefficients	t - value
Constant	42.042	2.015
Health organization	-0.177	-0.192
Retail organization	-1.915	-1.398
Manufacturing organization	0.817	1.157
other organization	0.099	0.262
Shifts	0.735	2.106
covered/uncovered bike parking, lockers a& showers, shelters, carpool & vanpool spaces, other amenities	-0.825	-1.608
Guaranteed ride home program	0.620	1.597
flexible timing	0.267	0.745
CTR Events, CTR E-mail, Articles, With Paychecks, Other Promo, etc	-0.093	-0.945
Ride match Apps, employees in-house match, employees public match	-0.303	-0.910
Ferry, Carpool, Walking, Bike Subsidy	0.282	0.685
SOV Parking Charge, SOV Parking Charge Number, Reduced SOVP Number	-0.346	-1.038
compressed work week 3/36, 4/40, 9/80, 7/40, other	-0.577	-0.733
Onsite incentives	0.584	1.315
Drawing, leaves, etc	1.045	2.578
Transit, vanpool subsidy	-0.101	-0.271
Number of employees	0.000	-0.012
Percentages of employees on 3/36 CWW	5.214	0.420
Percentages of employees on 4/40 CWW	-1.812	-0.418
Percentages of employees on 5/40	0.318	0.158
Percentages of employees on 7/40 CWW	-3.156	-0.394
Percentages of employees on 9/80 CWW	-1.915	-0.602
Percentages of employees on other CWW	-13.141	-1.944
Bike share	-14.208	-0.671
Bus share	-40.968	-2.042
Cars hare	-11.967	-1.050
Other share	-22.955	-0.983
Tele share	-53.781	-2.064
Van share	-15.433	-0.689
Walk share	-43.659	-2.108
VTR	-0.449	-2.198
Central Business District	-2.240	-0.567
Suburban area	-1.297	-0.331
Outside suburban area	-0.870	-0.223

Phase II: Washington Over-Sampled Data

To get better accuracies on moderate range bins a2 to a5, the examples in these bins were over-sampled. The Table 34 shows the changes in the number of examples in the training set due to over-sampling.

Table 32: Washington Over-Sampled Data – No. of Records in bins

Range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)							
		> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to -1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=
Bin Number		a1	a2	A3	a4	a5	a6	a7	a8
Validation	144	15	15	15	17	20	20	21	21
Training	1862	150	172	282	341	306	214	178	219

Neural network models were built of data containing the entire variable set (105 variables) with different variable selections. As discussed in the previous phase, the dataset contained variables such as shares of employee's incentive preference and other information that employers might not be able to provide to about their employees. These variables with a few other insignificant variables were removed from the data.

Then a neural network model with no variable selection and all different types of regression models were built on data without the preference variables. Figures 50, 51, and 52 show the different performance measure for all these models.

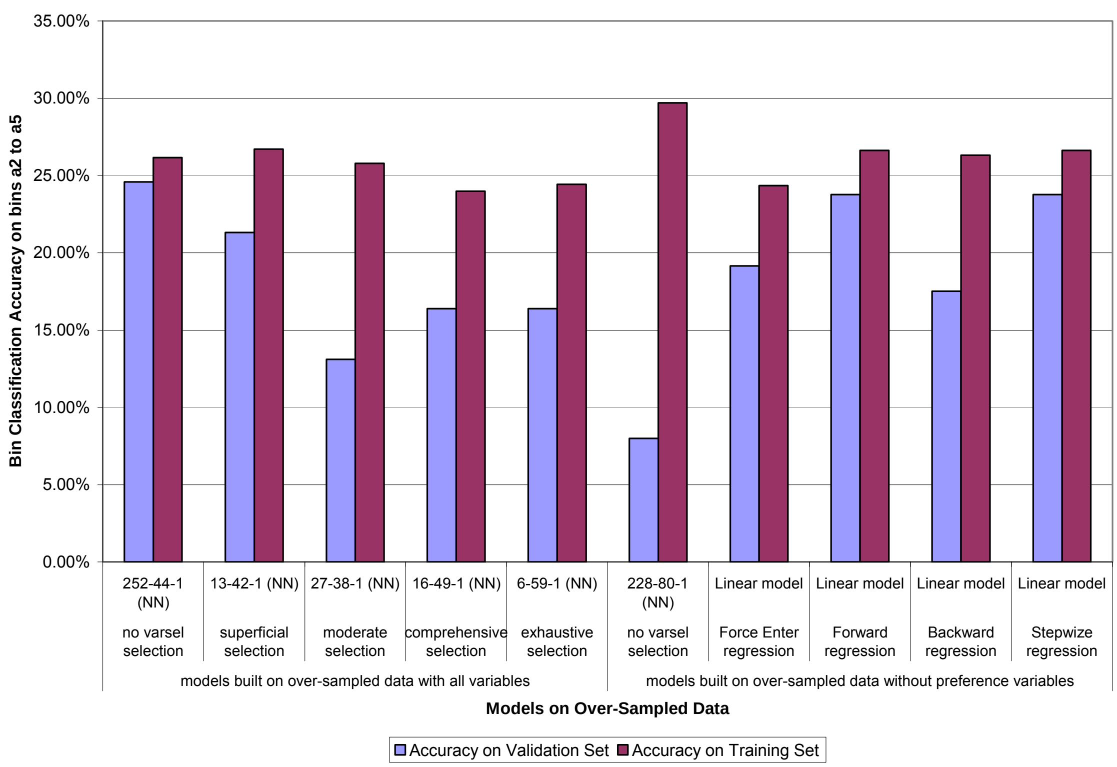


Figure 50: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on over-sampled data)

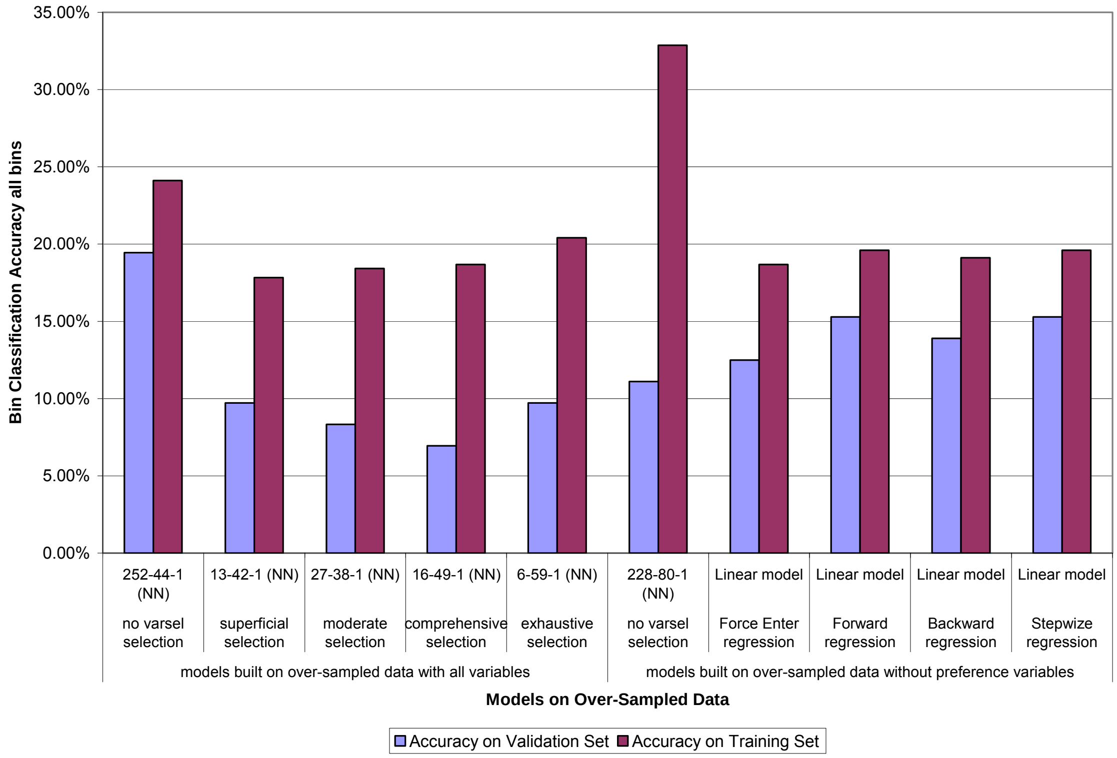


Figure 51: Bin Classification Accuracy on Full Range of change in VTR (all bins) for validation & training set (Different models on over-sampled data)

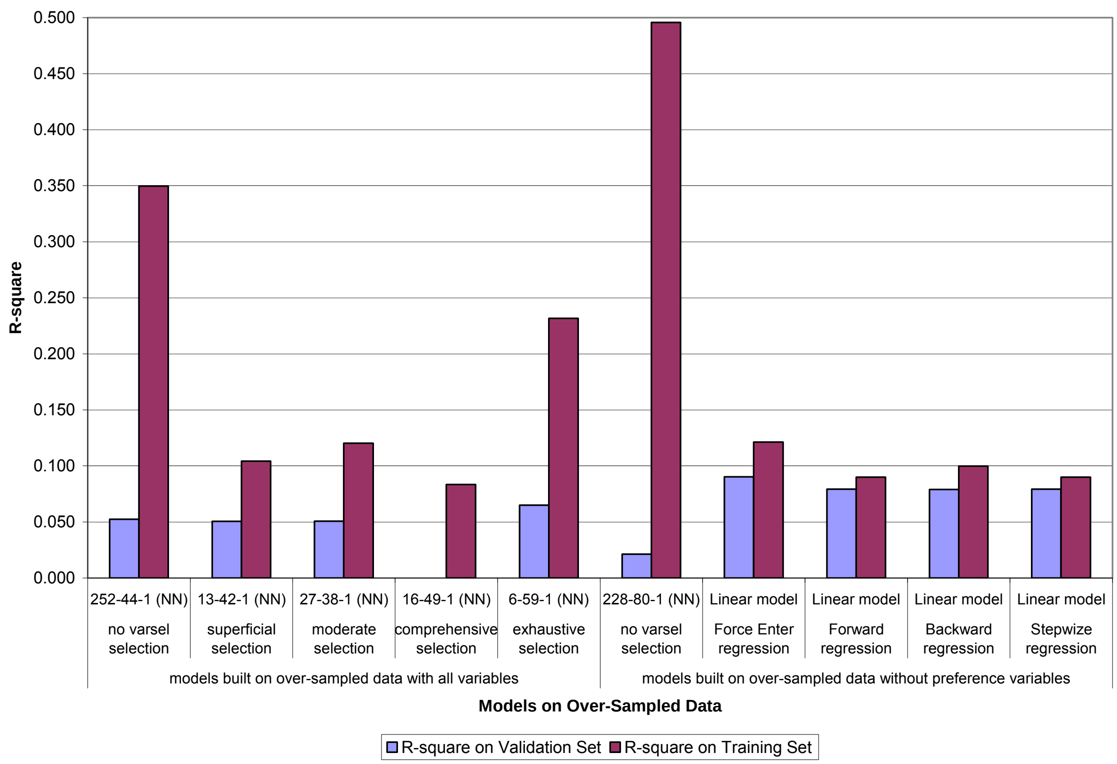


Figure 52: R-square for validation & training set (Different models on over-sampled data)

Table 35 shows the variables selected using the without preferences datasets to build models.

Column name abbreviations,

- N – Neural network model without variable selection
- S – Neural network model with superficial variable selection
- M – Neural network model with moderate variable selection
- C – Neural network model with comprehensive variable selection
- E – Neural network model with exhaustive variable selection
- FE – Linear Forced Enter regression model
- FR – Linear Forward regression model
- BK – Linear Backward regression model
- SP – Linear Stepwise regression model
- √ - indicates the selection of the variable
- '-' - indicates the variable was not present in the data

Table 33: Variables selected by different models on over-sampled data

Variables	NN models on data with all variables					Models on data without preference and other insignificant variables				
	N	S	M	C	E	N	FE	FR	BK	SP
Non-profit organization	√					-	-	-	-	-
Agriculture organization						-	-	-	-	-
Finance organization	√					-	-	-	-	-
Info services organization	√		√			-	-	-	-	-
Health organization	√					√	√			
Retail organization	√					√	√			
Manufacturing organization	√					√	√		√	
Services organization	√		√			-	-	-	-	-
Public utilities organization	√					-	-	-	-	-
Construction organization						-	-	-	-	-
Transportation organization	√					-	-	-	-	-
Government organization	√					-	-	-	-	-
Other organization	√					√	√		√	
Offered to All	√					-	-	-	-	-
Union	√					√	√			
Shifts	√		√			√	√		√	
Onsite Parking Spaces	√			√		√	√			
Offsite Parking Spaces	√		√			√	√		√	
Leased Onsite Parking Price	√					√	√			
Leased Offsite Parking Price	√			√		-	-	-	-	-
Own Onsite Parking	√			√	√	√	√	√	√	√
Onsite Parking Charge	√					√	√			
Own Offsite Parking	√					√	√			
Offsite Parking Charge	√					√	√			
Pay Parking Charge	√					√	√			
On off parking sub	√		√			√	√	√	√	√
Free Parking 1/4 mile	√					√	√		√	
ETC Additional Training	√					√	√			
ETC Onsite	√					√	√			
Distribute Info	√					√	√			
Post Materials	√					√	√			
CTR Orientation	√		√			√	√			
CTR Events	√					√	√	√	√	√
CTR E-mail	√			√		√	√			
Articles	√		√		√	√	√			
Ride match Apps	√		√			√	√			
With Paychecks	√					√	√			
Drawings	√		√	√	√	√	√	√	√	√
Leave						-	-	-	-	-
Other Promo	√			√		√	√			
Covered Bike Number	√					√	√			
Uncovered Bike Number	√			√		√	√			
Lockers Number	√					√	√			

Variables	NN models on data with all variables					Models on data without preference and other insignificant variables				
	N	S	M	C	E	N	FE	FR	BK	SP
Showers Number	√		√			√	√			
Shelters Number	√					√	√			
Other Amenities 1 Number	√				√	√	√	√	√	√
Carpool Spaces Number	√					√	√			
Vanpool Spaces Number	√					√	√			
SOV Parking Charge	√					√	√		√	
SOV Parking Charge Number	√		√			√	√			
Reduced SOVP Number	√	√				√	√		√	
Transit Subsidy	√					√	√			
Ferry Subsidy	√				√	√	√			
Vanpool Subsidy	√					√	√			
Carpool Subsidy	√					√	√			
Walking Subsidy	√					√	√			
Bike Subsidy	√					√	√			
Employees on flextime	√					√	√			
Employees with GRH	√				√	√	√			
Employees in-house match	√		√			√	√			
Employees public match	√			√	√	√	√			
FV work employees	√					√	√			
Number of Employees	√			√		√	√			
Percentages of employees on 3/36 CWW	√		√	√		√	√	√	√	√
Percentages of employees on 4/40 CWW	√	√	√	√		√	√	√		√
Percentages of employees on 5/40	√	√	√			√	√			
Percentages of employees on 7/40 CWW	√					√	√	√		√
Percentages of employees on 9/80 CWW	√			√		√	√			
Percentages of employees on other CWW	√					√	√	√	√	√
Alone share	√	√		√		√				
Bike share	√			√		√	√	√		√
Bus share	√	√	√	√	√	√	√		√	
Cars hare	√	√	√	√	√	√	√	√		√
Other share	√	√	√			√	√	√		√
Tele share	√					√	√			
Van share	√					√	√	√		√
Walk share	√			√		√	√		√	
Days saved telecommuting in two weeks	√					-	-	-	-	-
Administration job Share	√	√	√	√		-	-	-	-	-
Craft/Production/Labor Share	√			√		-	-	-	-	-
Management job Share	√					-	-	-	-	-
Sales/Marketing job Share	√	√				-	-	-	-	-
Customer Service job Share	√					-	-	-	-	-
Other job Share	√				√	-	-	-	-	-
Professional/Technical job Share	√					-	-	-	-	-
Q8_Invalid_Share	√		√			-	-	-	-	-
Prefer provide car for work	√					-	-	-	-	-
Employee Prefer Transport during lunch	√					-	-	-	-	-
Employee Prefer GRH	√					-	-	-	-	-
Employee Prefer flex to meet CVpool bus	√				√	-	-	-	-	-
Employee Prefer financial incentive	√			√		-	-	-	-	-
Employee Prefer reserved discounted CVpool space	√	√	√		√	-	-	-	-	-
Employee Prefer Personalized help for CVpool	√				√	-	-	-	-	-
Employee Prefer covered bicycle parking	√	√	√			-	-	-	-	-
Employee Prefer lockers & showers	√					-	-	-	-	-
Employee Prefer onsite childcare	√					-	-	-	-	-
Employee Prefer CWW	√					-	-	-	-	-
Employee Prefer to telecommute	√			√		-	-	-	-	-
Employee Prefer improved access to transit	√		√		√	-	-	-	-	-
SOV	√	√	√	√		-	-	-	-	-
VMT	√					-	-	-	-	-
VTR	√	√		√	√	√	√		√	

Variables	NN models on data with all variables					Models on data without preference and other insignificant variables				
	N	S	M	C	E	N	FE	FR	BK	SP
Central Business District	√					√	√			
Suburban area	√					√	√			
Outside suburban area	√					√	√			

From Figures 50 and 51, it can be seen that the neural network model built with no variable selection on data with all variables was able to get the best 'bin classification accuracy in moderate range of change in VTR' (24.59 percent) and 'best bin classification accuracy on full range of change in VTR' (19.44 percent) with an R-square value of 0.052. Since this model includes the preference variables, it again shows the importance of the employee's preferences towards incentives in predicting the changes in VTR.

The better models, excluding the previous model, are the forward/stepwise regression (M1) and forced enter regression (M2) models on the condensed data. The forward and stepwise regression models selected the same variables and regression coefficients, resulting in identical results. The detailed accuracies of these models on the each of the bins are shown in Table 36.

Table 34: Detailed accuracies on bins

	Range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
			> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to - 1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=		
	Bin Number		**a1**	**a2**	**a3**	**A4**	**a5**	**a6**	**a7**	**a8**		
	Validation	**144**	15	15	15	17	20	20	21	21	Weighted Avg on a2 to a5	R-square
	Training	**1862**	150	172	282	341	306	214	178	219		
M1	**Exact Validation**	15.28%	0.00%	0.00%	0.00%	50.00%	43.75%	15.38%	0.00%	13.64%	23.77%	0.079
	Exact Training	19.60%	0.67%	0.00%	12.77%	48.97%	40.85%	8.88%	5.62%	3.20%	26.63%	0.090
	One-off Validation	35.42%	0.00%	17.65%	0.00%	87.50%	87.50%	42.31%	5.56%	13.64%	49.39%	
	One-Off Training	51.61%	1.33%	12.21%	58.87%	93.84%	90.52%	50.47%	20.79%	13.70%	23.77%	0.079
M2	**Exact Validation**	12.50%	5.88%	5.88%	0.00%	31.25%	37.50%	11.54%	0.00%	4.55%	19.16%	0.090
	Exact Training	18.69%	1.33%	4.65%	13.12%	41.35%	35.29%	14.49%	7.87%	3.20%	24.36%	0.121
	One-off Validation	38.89%	11.76%	23.53%	0.00%	81.25%	81.25%	46.15%	5.56%	27.27%	47.85%	
	One-Off Training	52.95%	6.00%	25.00%	60.64%	91.50%	82.68%	50.93%	28.09%	17.81%		

It can be seen from the table, that for forwards forward/stepwise regressions models the accuracy of bins a1, a2, a3 and a7 is zero. This makes the model unsuitable as we would like to have accuracy spread out over all the bins.

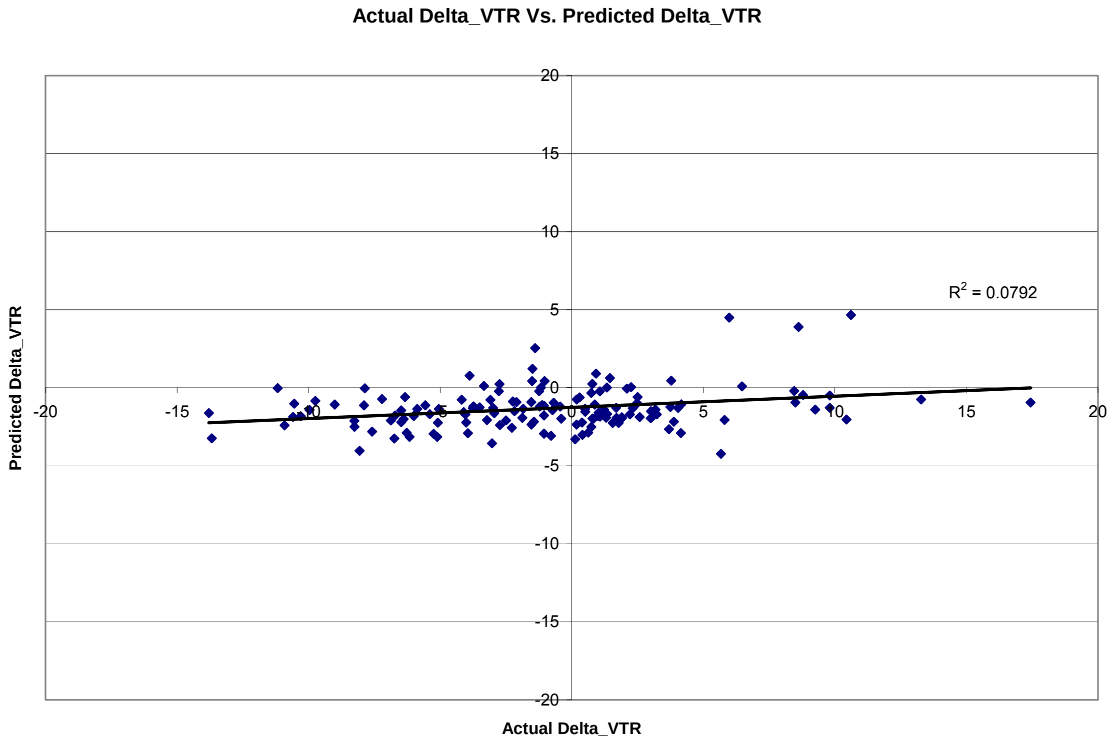


Figure 53 : Scatter plot for forward/stepwise regression model

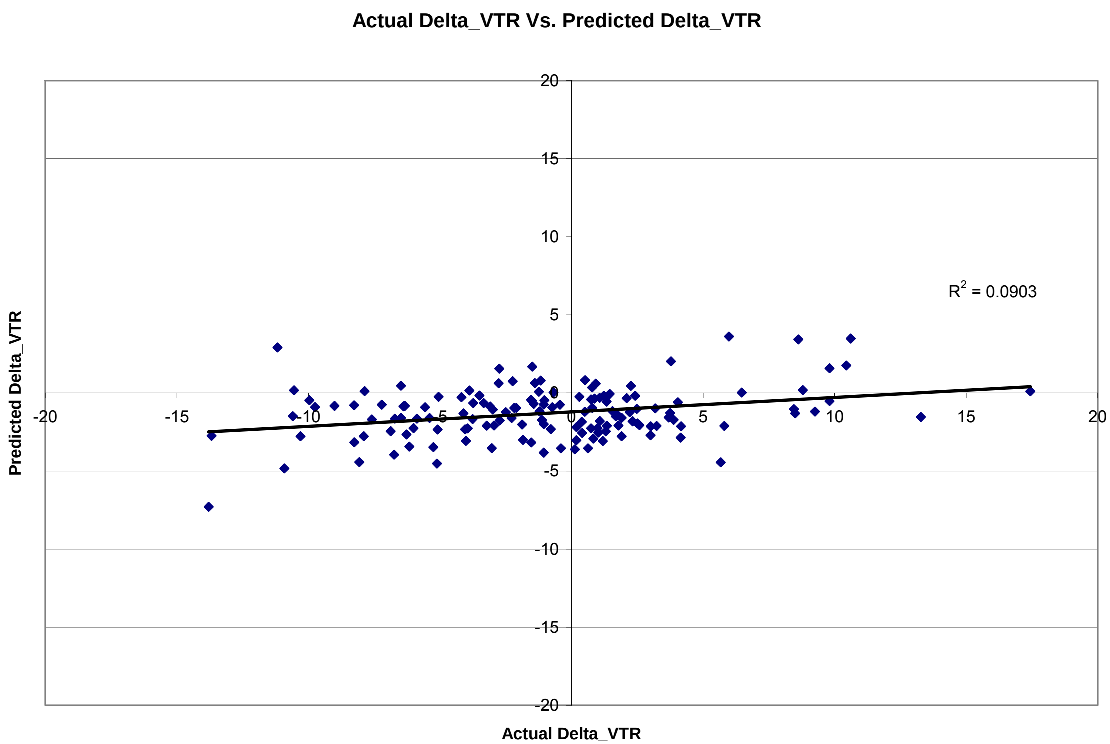


Figure 54: Scatter plot for forced enter regression model

The regression coefficients for the forced enter regression model (M2) are shown in Table 37.

Table 35: Variables and coefficients for forced enter regression model

Variable	coefficient	t - value
Constant	33.155	2.177
Health organization	-0.359	-0.511
Retail organization	-1.367	-1.422
Manufacturing organization	1.026	1.908
other organization	0.459	1.479
Shifts	0.521	1.943
Onsite Parking Spaces	-0.067	-0.473
Offsite Parking Spaces	-0.787	-1.805
Leased Onsite Parking Price	-0.001	-0.371
Leased Offsite Parking Price	-0.005	-1.206
Own Onsite Parking	0.791	2.886
Onsite Parking Charge	-0.001	-0.168
Own Offsite Parking	-0.412	-0.844
Offsite Parking Charge	0.006	0.830
Pay Parking Charge	0.006	1.453
on off parking sub	-1.519	-2.835
Free Parking 1/4 mile	0.368	1.403
ETC Additional Training	0.110	0.444
ETC Onsite	-0.082	-0.266
Distribute Info	0.004	1.344
Post Materials	0.001	0.437
CTR Orientation	0.000	0.763
CTR Events	-0.122	-3.177
CTR E-mail	-0.005	-0.718
Articles	0.020	1.049
Ride match Apps	-0.002	-1.187
With Paychecks	0.011	0.324
Drawings	0.050	2.062
Other Promo	0.005	1.017
Covered Bike Number	0.264	0.232
Uncovered Bike Number	2.908	0.892
Lockers Number	-0.068	-0.181
Showers Number	-5.687	-1.536
Shelters Number	-1.026	-0.315
Other Amenities 1 Number	-22.516	-3.710
Carpool Spaces Number	-1.373	-1.608
Vanpool Spaces Number	2.920	0.916
SOV Parking Charge	-0.009	-1.938
SOV Parking Charge Number	0.443	1.490
Reduced SOVP Number	2.798	1.652
Transit Subsidy	0.000	-0.003
Ferry Subsidy	0.014	1.410
Vanpool Subsidy	-0.022	-2.242
Carpool Subsidy	0.001	0.066
Walking Subsidy	0.014	0.468
Bike Subsidy	0.002	0.068
employees on flextime	0.350	0.686
employees with GRH	0.324	0.917
employees in-house match	-0.191	-1.328
employees public match	-0.113	-1.345
FV work employees	-0.043	-0.340
Number of Employee	0.000	-0.517
Percentages of employees on 3/36 CWW	7.452	0.799
Percentages of employees on 4/40 CWW	-2.340	-0.731
Percentages of employees on 5/40	-1.600	-1.029
Percentages of employees on 7/40 CWW	-3.723	-0.651

Variable	coefficient	t - value
Percentages of employees on 9/80 CWW	-4.338	-1.842
Percentages of employees on other CWW	-15.095	-3.142
Bike share	-6.940	-0.440
Bus share	-30.244	-2.070
Cars hare	-7.566	-0.909
Other share	-14.234	-0.819
Tele share	-27.542	-1.443
Van share	-6.184	-0.376
Walk share	-35.160	-2.316
VTR	-0.334	-2.252
Central Business District	-3.219	-1.154
Suburban area	-2.928	-1.059
Outside suburban area	-2.520	-0.914

When the scatter plot in figure 54 for this model is compared with the scatter plot in figure 43 for equivalent forced enter regression model built on the full sampled ungrouped incentive data, we can see that the model on over-sampled data is biased towards predicting negative changes in VTR.

Equivalent grouped incentives models were also built on the over-sampled data. Figures 55, 56, and 57 show the accuracies and the R-square values of the neural network and the regression models built using no variable selection.

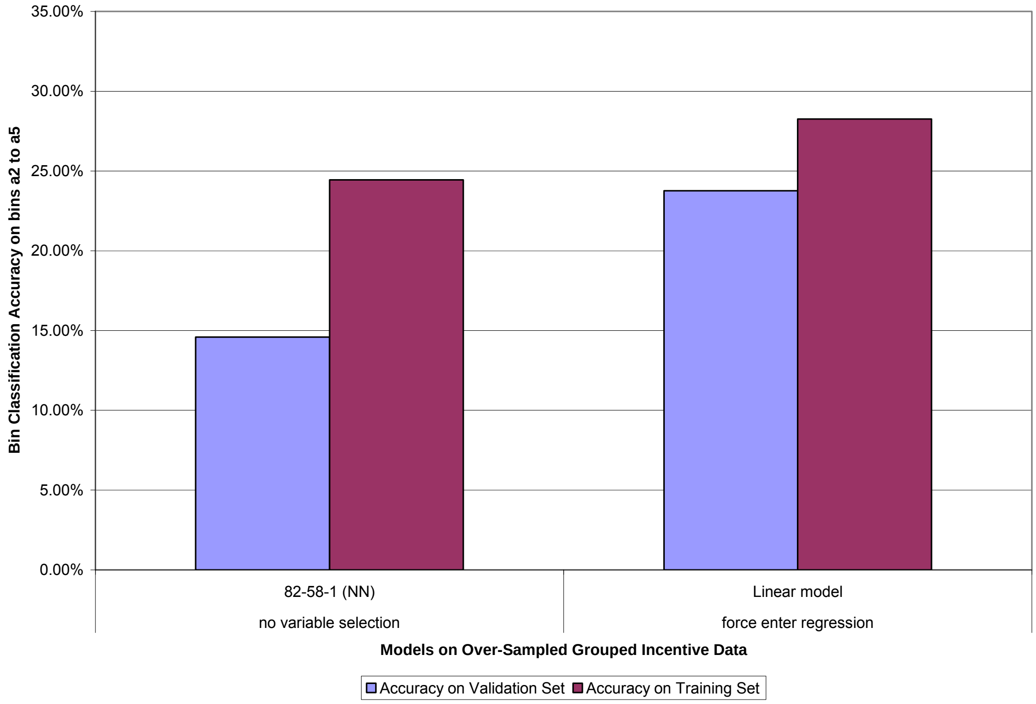


Figure 55: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for validation & training set (Different models on Over-Sampled Grouped Incentive data)

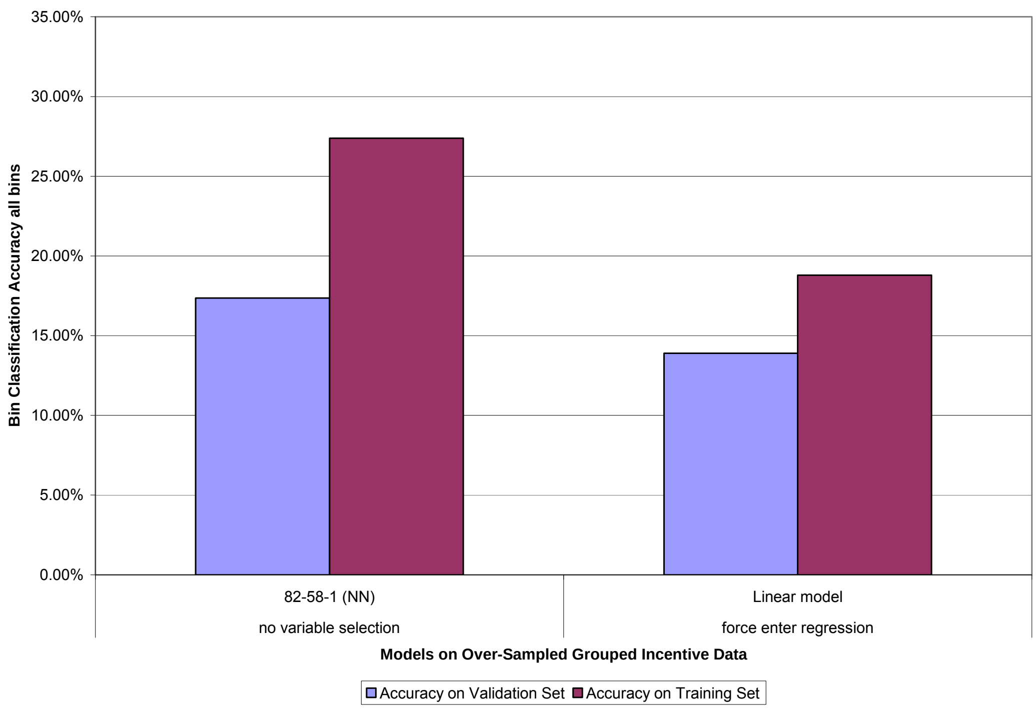


Figure 56: Bin Classification Accuracy on Full Range of change in VTR (all bins) for validation & training set (Different models on Over-Sampled Grouped Incentive data)

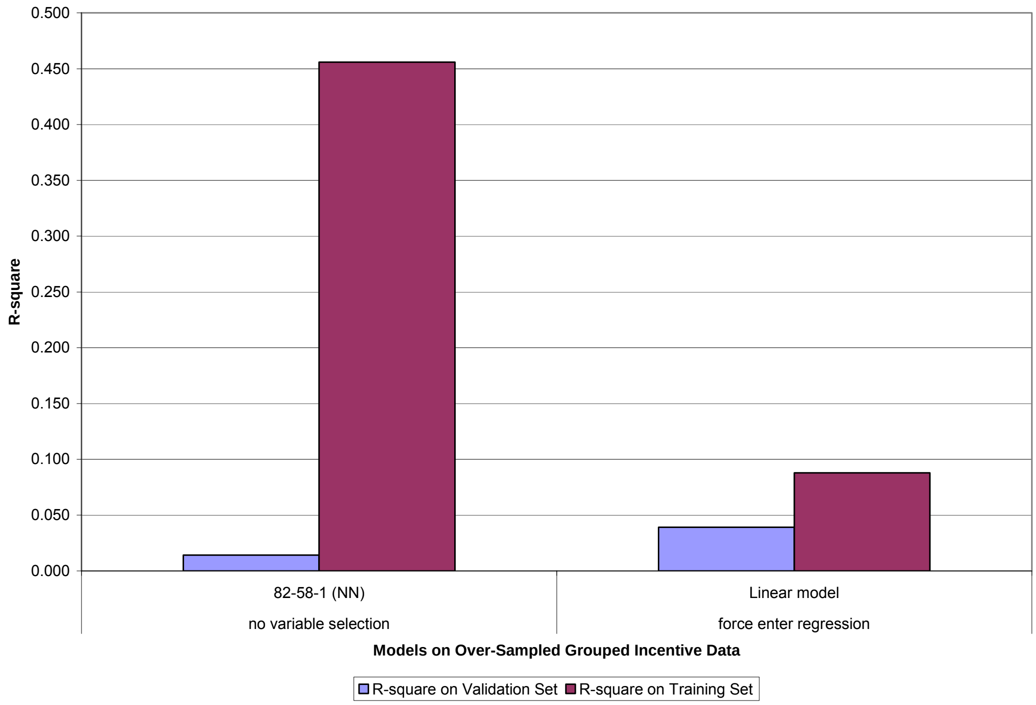


Figure 57: R-square for validation & training set (Different models on Over-Sampled Grouped Incentive data)

The charts in Figures 55, 56, and 57 display that none of the models beat the others in all of the performance measures. The Neural network (M1) and linear regression (M2) built on this over-sampled data seem to do better than the corresponding models built on non-over-sampled data. The accuracies for both of the models are shown in Table 38.

Table 36: Detailed accuracies on bins

	Range	# of records	Bin ranges over Change in Vehicle Trip Rate (Delta_VTR)									
			> - 7	[- 7 to - 4.5)	[- 4.5 to - 3)	[- 3 to - 1.5)	[-1.5 to 0)	[0 to 1.5)	[1.5 to 3.5)	3.5 >=		
	Bin Number		A1	a2	a3	a4	a5	a6	A7	a8		
	Validation	144	15	15	15	17	20	20	21	21	Weighted Avg on a2 to a5	R-square
	Training	1862	150	172	282	341	306	214	178	219		
M1	Exact Validation	17.36%	0.00%	0.00%	8.33%	18.75%	31.25%	19.23%	27.78%	27.27%	14.58%	0.014
	Exact Training	27.39%	15.33%	13.29%	19.78%	26.05%	33.23%	23.26%	27.78%	53.88%	24.45%	0.456
	One-off Validation	37.50%	5.88%	11.76%	33.33%	50.00%	50.00%	38.46%	50.00%	54.55%	35.79%	
	One-Off Training	63.86%	38.67%	40.46%	58.63%	64.67%	76.36%	70.23%	68.89%	76.71%	14.58%	0.010
M2	Exact Validation	13.89%	0.00%	0.00%	0.00%	50.00%	43.75%	11.54%	0.00%	9.09%	23.77%	0.039
	Exact Training	18.80%	0.00%	0.00%	8.99%	49.70%	38.02%	9.77%	7.78%	2.28%	28.26%	0.088
	One-off Validation	34.72%	0.00%	0.00%	50.00%	87.50%	87.50%	46.15%	5.56%	13.64%	55.12%	
	One-Off Training	50.43%	2.00%	18.50%	54.68%	95.21%	83.39%	50.23%	21.67%	11.87%		

The scatter plots on the validation set for these models built on over-sampled data are shown in Figures 58 and 59.

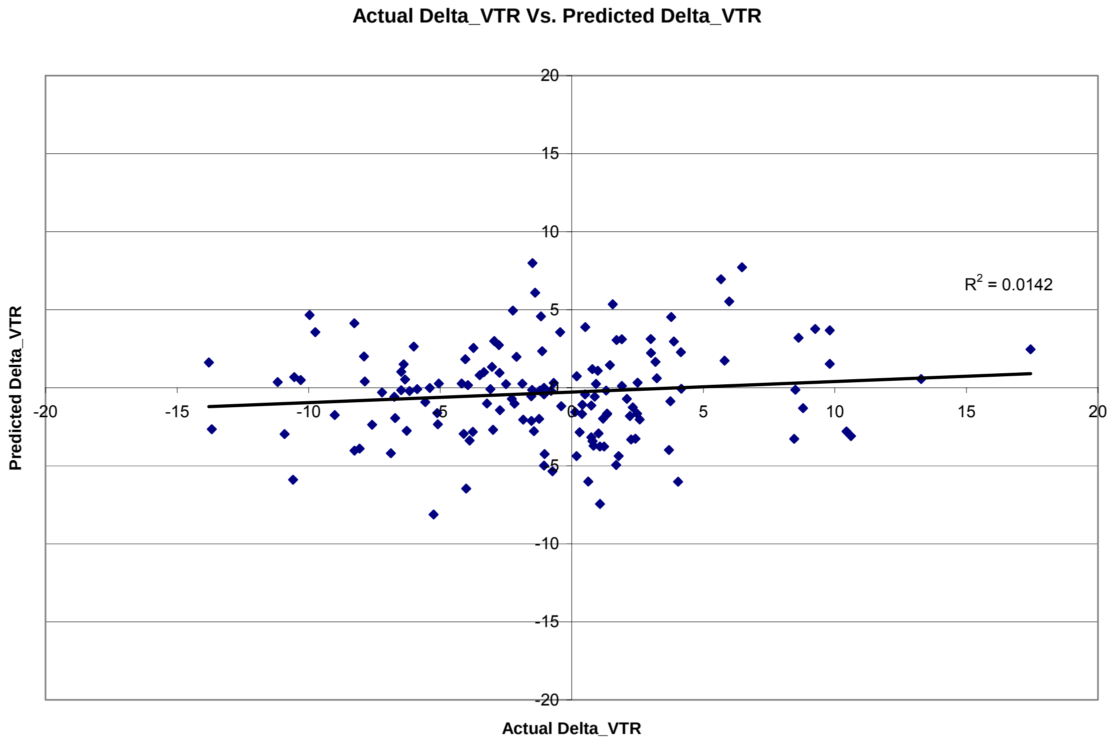


Figure 58: Scatter plot for neural network model

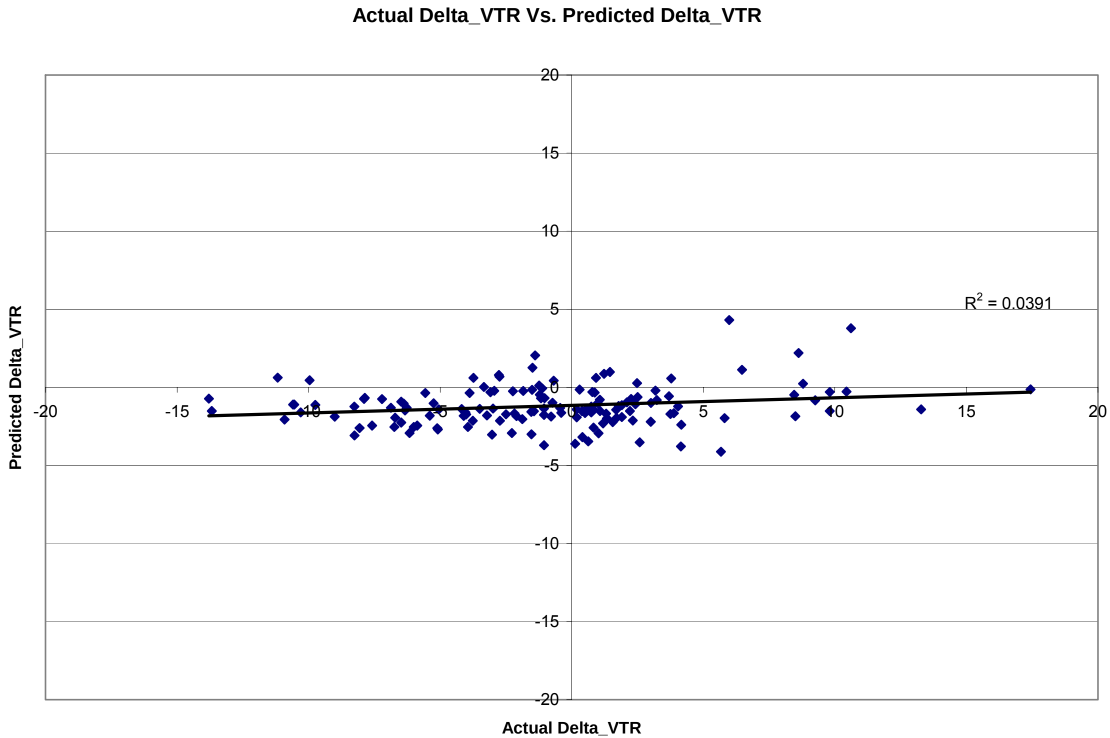


Figure 59: Scatter plot for linear regression model

From the detailed accuracy table 38 and scatter plot in figure 58, it can be seen that the linear forced enter regression model is biased towards predicting negative changes in VTR which is not the expected behavior of a good model.

The regression coefficients for the forced enter regression model (M2) are shown in Table 39.

Table 37: Variables and coefficients for forced enter regression model

Variables	coefficients	t - value
(Constant)	22.871	1.548
Health organization	-0.234	-0.344
Retail organization	-1.589	-1.660
Manufacturing organization	0.411	0.817
other organization	0.023	0.083
Shifts	0.695	2.724
covered/uncovered bike parking, lockers a& showers, shelters, carpool & vanpool spaces, other amenities	-0.650	-1.741
Guaranteed ride home program	0.617	2.179
flexible timing	0.128	0.497
CTR Events, CTR E-mail, Articles, With Paychecks, Other Promo, etc	-0.052	-0.725
Ridematch Apps, employees in-house match, employees public match	-0.299	-1.237
Ferry, Carpool, Walking, Bike Subsidy	0.074	0.245
SOV Parking Charge, SOV Parking Charge Number, Reduced SOVP Number	-0.319	-1.325
Compressed work week 3/36, 4/40, 9/80, 7/40, other	-0.278	-0.482
onsite incentives	0.503	1.552
Drawing, leaves, etc	0.619	2.125
Transit, vanpool subsidy	-0.038	-0.141
Number of employees	0.000	0.159
Percentages of employees on 3/36 CWW	13.872	1.521
Percentages of employees on 4/40 CWW	1.186	0.382
Percentages of employees on 5/40	0.225	0.157
Percentages of employees on 7/40 CWW	0.057	0.010
Percentages of employees on 9/80 CWW	-1.278	-0.566
Percentages of employees on other CWW	-10.221	-2.134
Bike share	-0.549	-0.036
Bus share	-22.824	-1.605
Cars hare	-3.658	-0.449
Other share	-4.527	-0.265
Tele share	-30.459	-1.622
Van share	0.494	0.031
Walk share	-25.494	-1.724
VTR	-0.253	-1.750
Central Business District	-2.317	-0.830
Suburban area	-1.693	-0.612
Outside suburban area	-1.355	-0.492

Recommended Model

The neural network models built in phase I and II on data containing the employee's preferences obtained much better accuracies than the ones that were built without them. This shows the importance of the employee's preferences towards incentives in predicting the changes in VTR. But as stated earlier, the employers may not have access to this employee information. Therefore, these models can only be used for study.

None of the models built without the employee's incentive preferences were exceptionally better from one another. The two models that stood out were the forced enter regression model built on full sample data and the neural network models built on the over-sampled grouped incentive data.

The rationale for considering the forced enter regression model built on full sample data as one of the candidate model was (Table 29-M1)

1. It has 'bin classification accuracy on moderate range of change in VTR' of 19.40%
2. It has a 'R-square' value of 0.047
3. It has 'bin classification accuracy on full range of change in VTR' of 15.97% which is better than the random choice accuracy of 12.5%
4. It includes all the individual incentives in predicting change in VTR

The rationale for considering the neural network models built on the over-sampled grouped incentive data as one of the candidate model was (Table 35-M1)

1. It has the best 'bin classification accuracy on full range of change in VTR' of 17.36% which is much better than the random choice accuracy of 12.5%
2. It has 'bin classification accuracy on moderate range of change in VTR'' of 14.58%
3. It has a 'R-square' value of 0.014
4. It is built on simple grouped incentives variable set.

It was very difficult to select one of these models as the best one. Accordingly, a cross-tab for positive/negative actual changes in VTR against positive/negative predicted changes in VTR on validation and training/testing set was done as shown in tables 40 and 41.

Table 38 A-D: Validation Set

A

Forced enter regression model on full sample incentive data		
Actual Delta_VTR / Predicted Delta_VTR	**Negative**	**Positive**
Negative	56 (TN)	22 (FP)
Positive	45 (FN)	21 (TP)

B

NN model on over-sampled grouped incentive data		
Actual Delta_VTR / Predicted Delta_VTR	**Negative**	**Positive**
Negative	44	34
Positive	38	28

C

Forced enter regression model on full sample incentive data	
Precision	**0.718**
Recall	**0.554**
F-value	**0.626**

D

NN model on over-sampled grouped incentive data	
Precision	0.564
Recall	0.537
F-value	0.550

Table 39 A-D: Training/testing Set

A

Forced enter regression model on full sample incentive data		
Actual Delta_VTR / Predicted Delta_VTR	**Negative**	**Positive**
Negative	603	135
Positive	355	207

B

NN model on over-sampled grouped incentive data		
Actual Delta_VTR / Predicted Delta_VTR	**Negative**	**Positive**
Negative	932	316
Positive	186	428

C

Forced enter regression model on full sample incentive data	
Precision	**0.817**
Recall	0.629
F-value	0.711

D

NN model on over-sampled grouped incentive data	
Precision	0.747
Recall	**0.834**
F-value	**0.788**

Where,
TN – True Negatives: These are the number of records for which the model predicted a negative change in VTR when the actual change in VTR was also negative (we are more interested in these)

TP – True Positives: These are the number of records for which the model predicted a positive change in VTR when the actual change in VTR was also positive

FP – False Positives: These are the number of records for which the model falsely predicted a positive change in VTR when the actual change in VTR was negative

FN – False Negatives: These are the number of records for which the model falsely predicted a negative change in VTR when the actual change in VTR was also positive

Since we are more interested in the accuracy of the model predicting negative changes in VTR (i.e. we modeling for predicting negative change in VTR), precision gives us the measure of correctness of the model in predicting negative changes in VTR.

$$\text{Precision} = \frac{\text{True negative (TN)}}{\text{True negative (TN)} + \text{False negative (FN)}}$$

Recall gives us the measure of completeness of the model in predicting negative changes in VTR

$$\text{Recall} = \frac{\text{True negative (TN)}}{\text{True negative (TN)} + \text{False positive(FP)}}$$

$$\text{F - value} = \frac{(1+\beta^2) \times \text{Recall} \times \text{Precision}}{\beta^2 \times \text{Recall} + \text{Precision}}$$

Where $\beta = 1$ (β controls relative importance between recall and precision)

Because the desire is to have better accuracies in predicting negative changes in VTR, it would be good to improve the recall without sacrificing the precision. Both of these measures are captured by the F-value and so the goodness of the model can be measured by F-value. It is apparent that the forced enter regression model built on full sample ungrouped incentives data was better than the neural net model built on the over-sampled grouped incentives data. This is because the forced enter regression model has obtained a better f-value on validation set whereas the neural network model got better f-value on training set but much less f-value on validation set which might be the due to over-fitting of the model on training set and losing its generalization power on validation set. Therefore, the forced enter regression model built on full sample data is the recommended model.

COMBINED DATA MODELS

The grouped incentive data from all the three datasets was combined to build a generalized model. Some of the variables were collapsed into one variable as shown in Table 42, so that all of the datasets had consistent variables.

Table 40: Variable mapping for combining all three datasets

Description	Combined Data	Tucson Grouped	LA Grouped	Wash Grouped
Alone Share	AloneShare	AloneShare	CAR1	AloneShr
Transit Share	TransitShare	BusShare	TRANSIT	BusShare
Carpool +Vanpool Share	CVpoolShare	CVpoolShare	VAN_CUTR + CAR2 + CAR3 + CAR4 + CAR5 + CAR6 + BUS	CpoolShr + VpoolShr
Walk Share	WalkShare	WalkShare	WALK	WalkShr
Bicycle Share	BcycleShare	McycleShare	BIKE	BikeShr
Motorcycle Share	McycleShare	BcycleShare	Motorcycle	OtherShr
Telecommute Share	TeleShare	0 (not present)	TELECOMMUTE	TeleShr
3/36 compressed work week share	CWW336	CWW336	CWW336	CWW336
4/40 compressed work week share	CWW440	CWW440	CWW440	CWW440
8/80 compressed work week share	CWW980	CWW980	CWW980	CWW980
facilities & amenities	FACILITY_AMENITIES	FACILITY_AMENITIES	FACILITY_AMENITIES	FACILITY_AMENITIES
Guaranteed ride home programs	GRH	GRH	GRH	GRH
flexible timing	FLEX	FLEX	FLEX	FLEX
Marketing programs	MRKT	MRKT	MRKT	MRKT
Ride share matching programs	RS_MATCH	RS_MATCH	RS_MATCH	RS_MATCH
financial incentives	FINANCIAL	FINANCIAL	FINANCIAL	FINANCIAL
Parking management	PARKMGT	PARKMGT	PARKMGT	PARKMGT
Telecommute program	TELE	TELE	TELE	TELE
Compressed work week program	CWW	CWW	CWW	CWW
onsite incentives	ONSITE	ONSITE	ONSITE	ONSITE
Non financial incentives	DIRECT_NONFINAN	DIRECT_NONFINAN	DIRECT_NONFINAN	DIRECT_NONFINAN
commuter tax benefit incentives	COMMTAX	COMMTAX	COMMTAX	COMMTAX
Vehicle trip rate	VTR	VTR	VTR	VTR

Phase I: Combined Data

The training data from all the three cities was combined into a single training set while the validation sets for each city were left undisturbed. The number of examples in each bin for three validation and training/testing sets are shown in Table 43.

Table 41: Combined Data – No. of Records in bins

Bin Number	Total	a1	a2	a3	a4	a5	a6	a7	a8
Validation (LA)	**1838**	292	220	261	290	173	157	186	259
Validation (Tucson)	**112**	5	4	9	17	17	21	19	20
Validation (wash)	**144**	17	17	13	15	16	26	18	22
Train +Test	**19173**	2729	2638	2732	3287	1713	1803	2009	2262

A simple neural network with no variable selection and forced enter linear regression model were built on the data just containing 23 variables shown in Table 42. The models were trained on the combined training set data from three cities and then the model was independently evaluated on the separate validation sets from these cities. These validation sets were the same validation sets used in evaluating the models built independently on each city data. Figures 60, 61 and 62 show the comparison of the bin classification accuracies and R-square values of the models built with the combined data and the recommended models built separately on the three datasets.

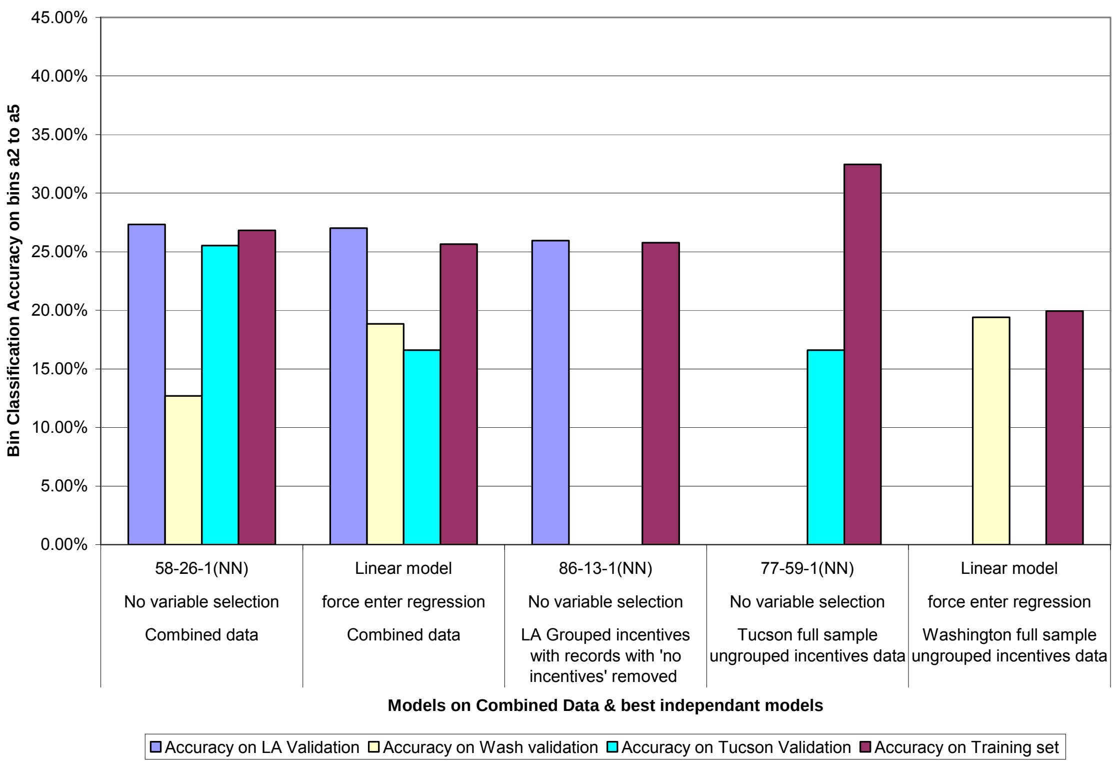


Figure 60: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for three data sets (Models on combined data & best independent models)

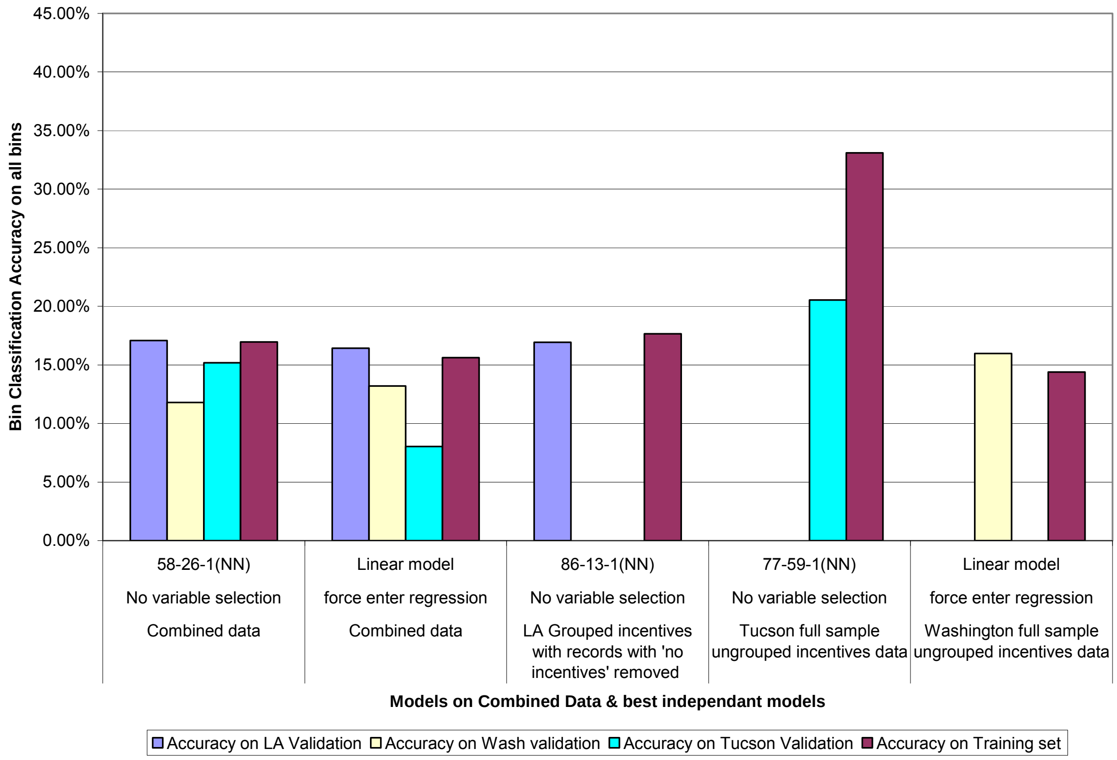


Figure 61: Bin Classification Accuracy on Full Range of change in VTR (all bins) for three validation & training sets (Models on combined data & best independent models)

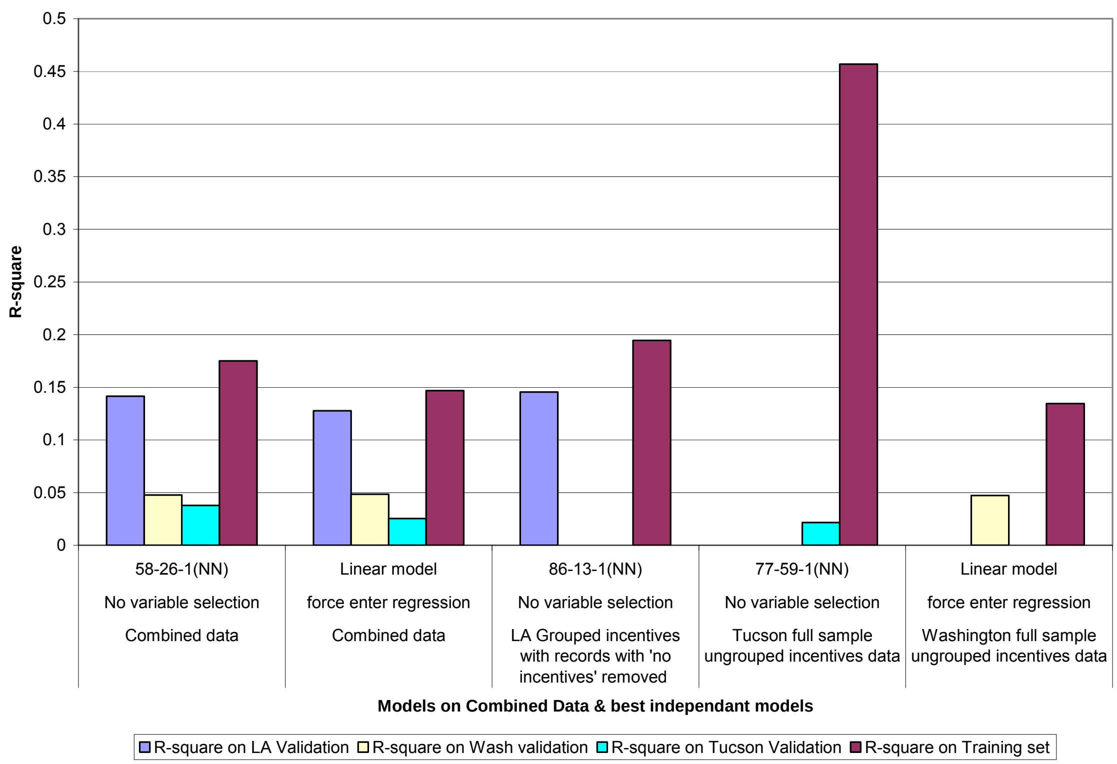


Figure 62: R-square for three validation & training set (Models on combined data & best independent models)

The results in Figure 60, show that the neural net model built on the combined data with no variable selection was able to get better 'bin classification accuracy on moderate range of change in VTR' on the Los Angeles validation set (27.33 percent) as compared to the recommended independent Los Angeles model (25.95 percent) built on grouped incentive dataset and also better 'bin classification accuracy on moderate range of change in VTR' on the Tucson validation set (25.53 percent) as compared to the recommended independent Tucson model (16.6 percent) built on full sample ungrouped incentive dataset. Neither of the models built with combined data were able to improve the accuracy of the Washington validation data.

From figure 61, it was found that this neural net model built on the combined data with no variable selection was able to obtain better 'bin classification accuracy on full range of change in VTR' on the Los Angeles validation set (17.08 percent) as compared to the recommended independent Los Angeles model (16.92 percent). There was considerable reduction in 'bin classification accuracy on full range of change in VTR' on the Tucson and Washington validation sets when compared to result got by independent recommended models.

Figure 62 shows that some improvement in R-square value was obtained by combined data neural net model on the Tucson validation set. The scatter plots for the neural net model built on the combined data of the three validation sets are shown in Figures 63, 64, and 65.

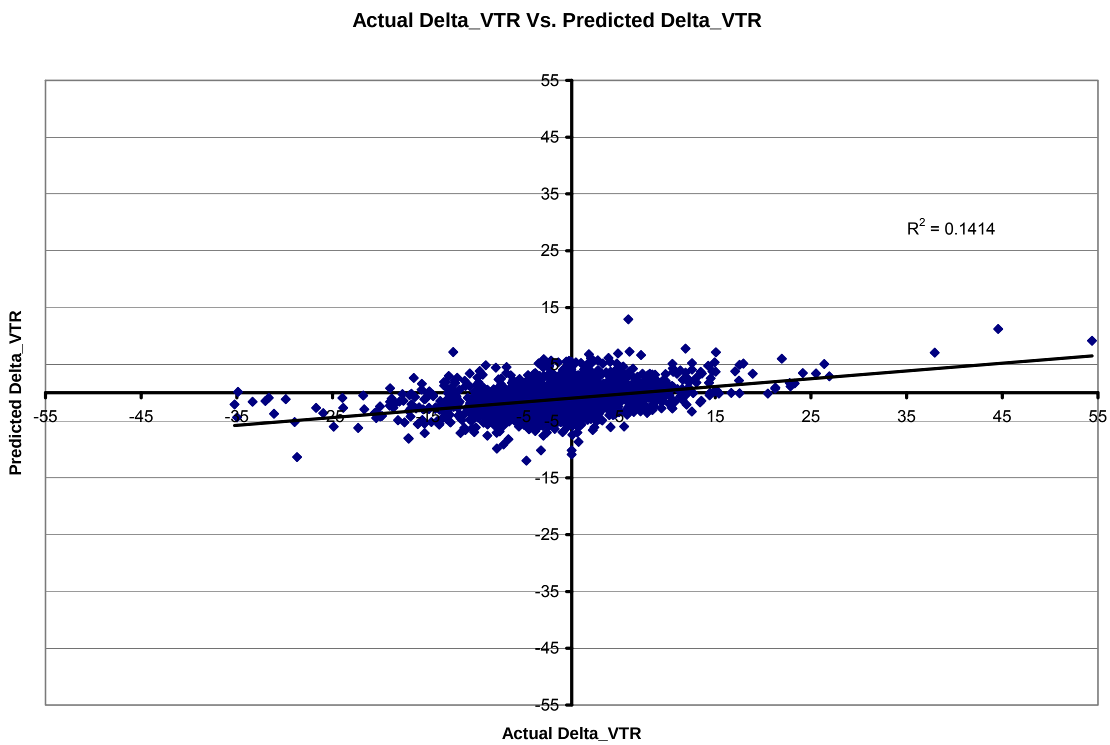


Figure 63: Scatter plot for Los Angeles validation set

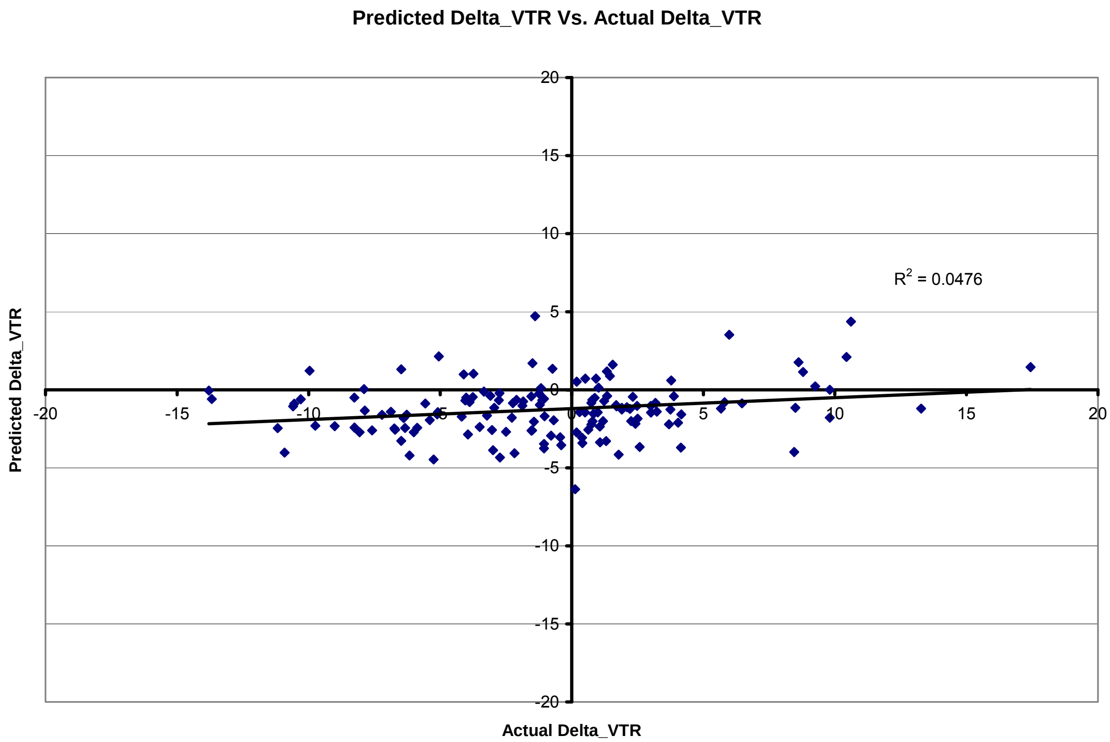


Figure 64: Scatter plot for Washington validation set

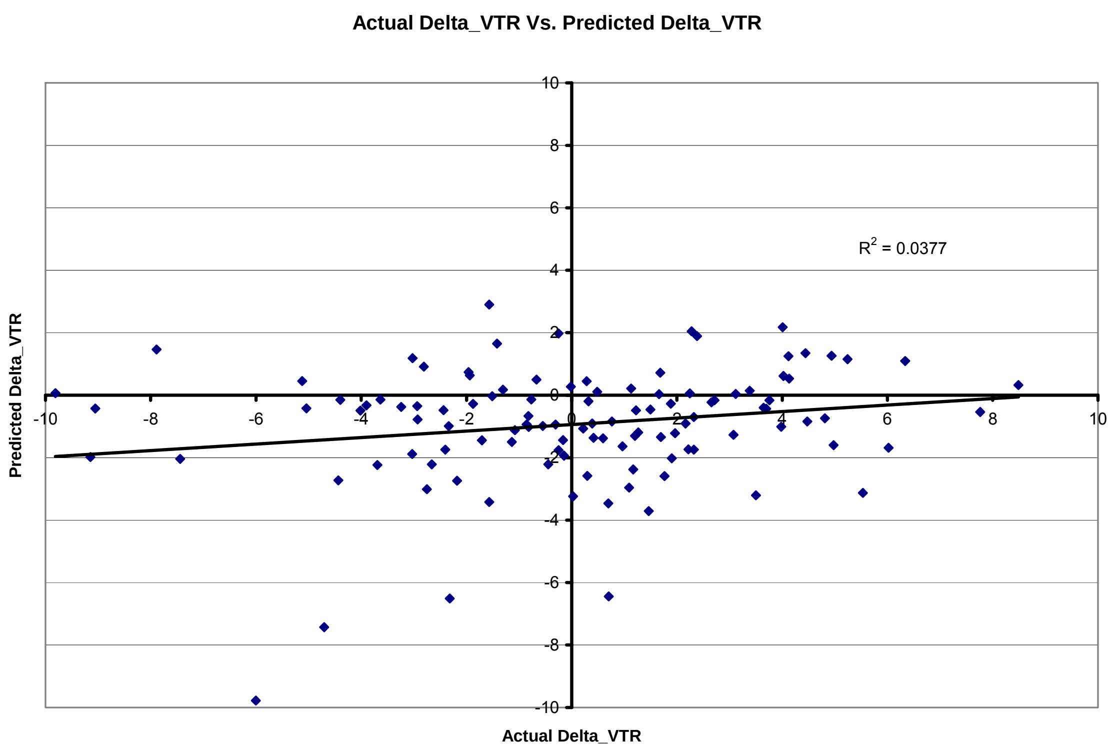


Figure 65: Scatter plot for Tucson validation set

It can be concluded from these results that adding more information from other cities helped in improving the accuracy for Los Angles data, and so the neural network model built on this data is the recommended model for Los Angeles data. For Tucson data, the accuracy charts did not reveal much about the goodness of the neural net model built on the combined data, but from the scatter plot for Tucson validation set it can be seen that the combined data model was much biased at predicting negative changes in VTR, rendering it unsuitable. For the Washington data, adding information from the other cities proved disadvantageous. This result might be due to a very large share of training records coming from the Los Angeles data.

Phase II: Equally Sampled Combined Data

To alleviate the problem that caused the unusual results for Tucson and Washington data on the non-equalized combined data, the training data from all of the three cities was combined in such a manner that an equal number (2,018) of records from each data set contributed to the combined training data. The number of examples in each bin for three validation and training/testing sets are shown in Table 44.

Table 42: Equalized Combined Data – No. of Records in bins

	Total	a1	A2	a3	a4	a5	a6	a7	a8
Validation (LA)	1838	292	220	261	290	173	157	186	259
Validation (Tucson)	112	5	4	9	17	17	21	19	20
Validation (wash)	144	17	17	13	15	16	26	18	22
Train +Test	6054	0	46	1371	3059	777	514	226	61

A simple neural network with no variable selection and a forced enter linear regression model were built with the equally sampled training data. This combined model only contained 23 variables, which are shown in Table 42. The charts in Figures 66, 67, and 68 show the results of the comparison between the three performance measures for the combined, equally sampled models.

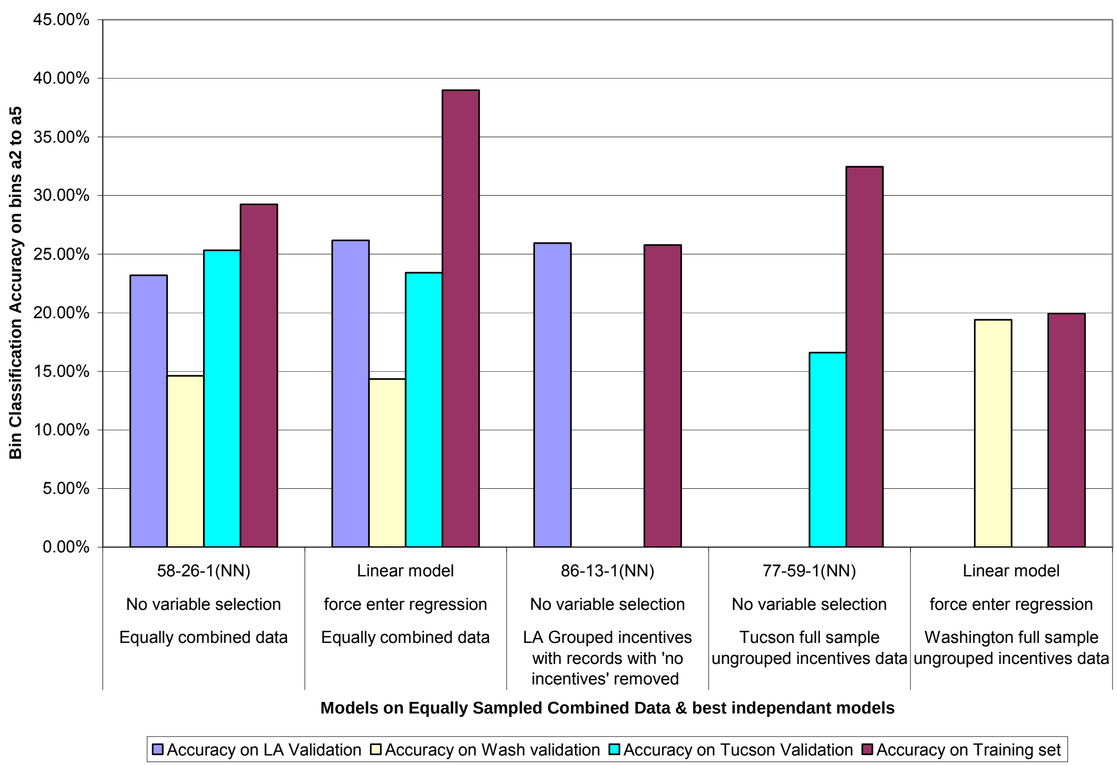


Figure 66: Bin Classification Accuracy on Moderate Range of change in VTR (a2 to a5) for three data sets (Models on equally combined data & best independent models)

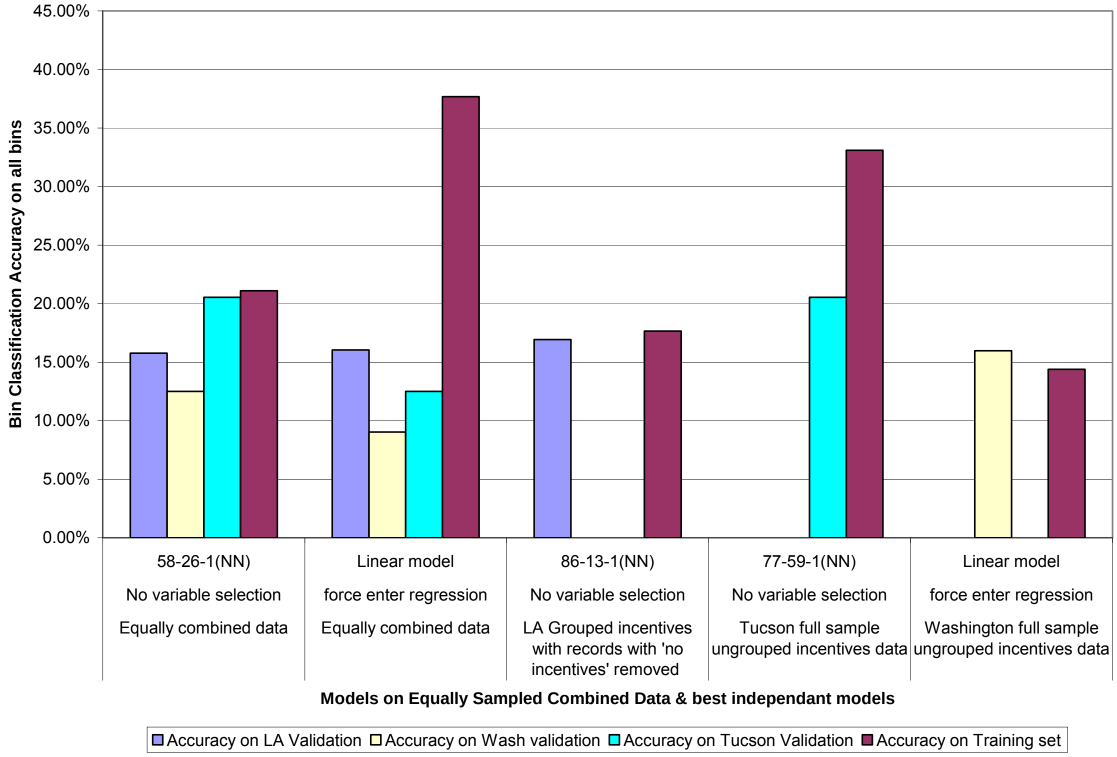


Figure 67: Bin Classification Accuracy on Full Range of change in VTR (all bins) for three validation & training sets (Models on equally combined data & best independent models)

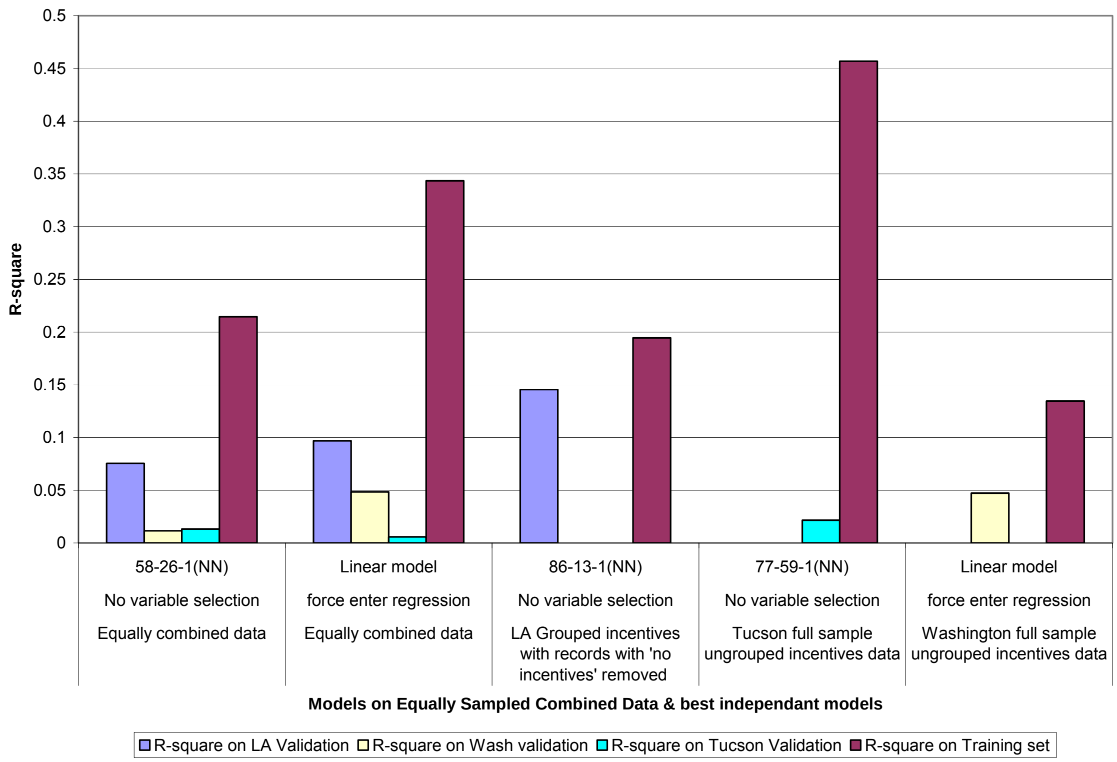


Figure 68: R-square for three validation & training set (Models on equally combined data & best independent models)

Figure 66 displays the comparison of the results between the neural net model with no variable selection and the forced enter regression model built on the equally sampled combined data and the recommended independent models on three datasets. Both combined data models were able to get better 'bin classification accuracy on moderate range of change in VTR' on Tucson validation set (NN model: 25.32 percent and Linear model: 23.40 percent) as compared to the recommended independent Tucson model (16.6 percent) but were not able to improve any accuracy on Los Angeles and Washington validation sets.

From Figure 67, it can be seen that neural net model built on the equally sampled combined data with no variable selection was able to get same 'bin classification accuracy on full range of change in VTR' on Tucson validation set (20.54 percent) as the recommended independent Tucson model. No other significant improvements were obtained by these combined data models over other recommended independent models. The same was true for R-squares values.

The scatter plots for the neural net model built on combined data on three validation sets are shown in Figures 69, 70, and 71.

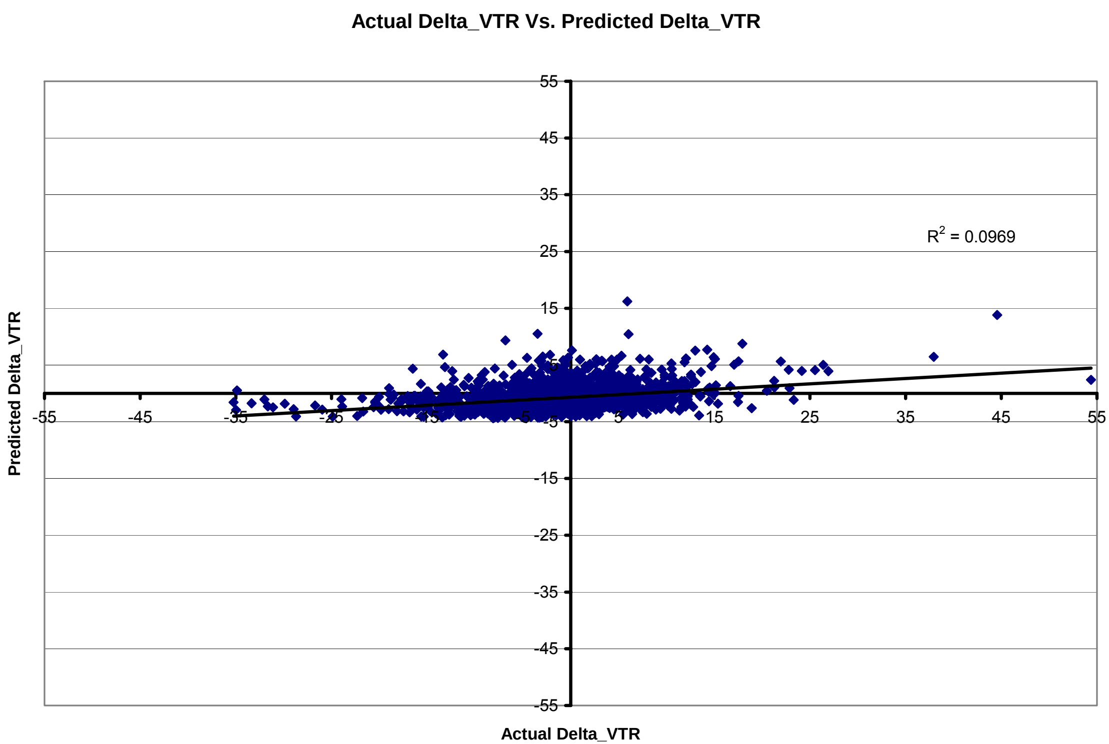


Figure 69: Scatter plot for Los Angeles validation set

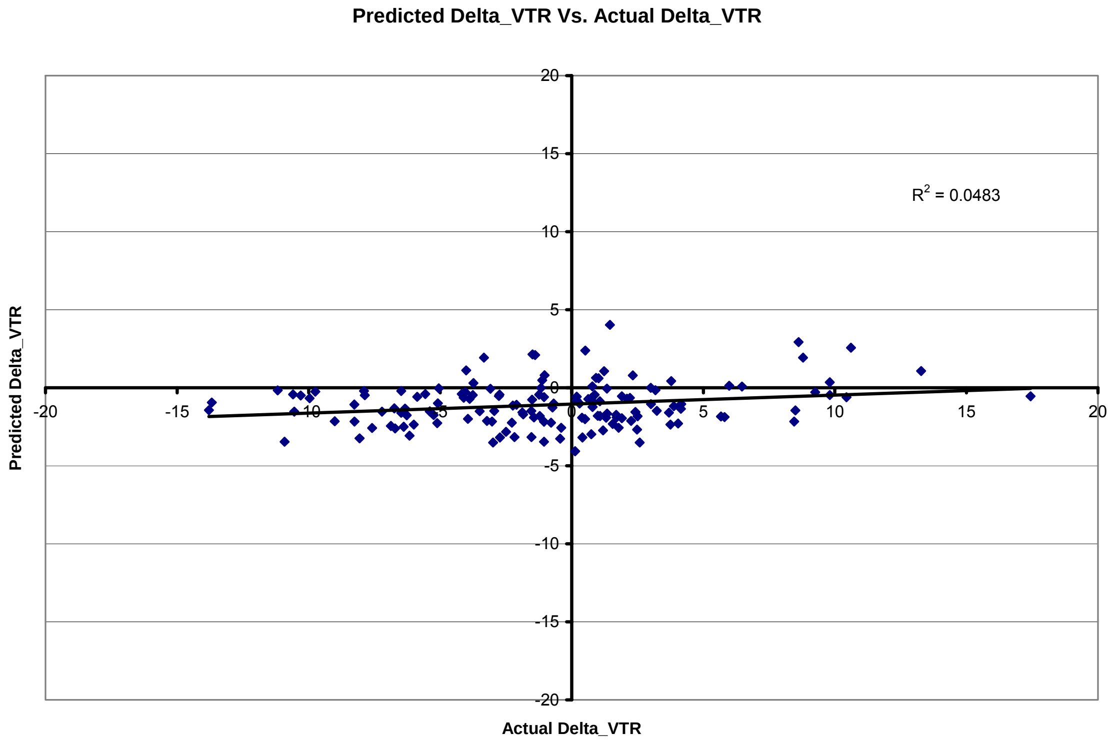


Figure 70: Scatter plot for Washington validation set

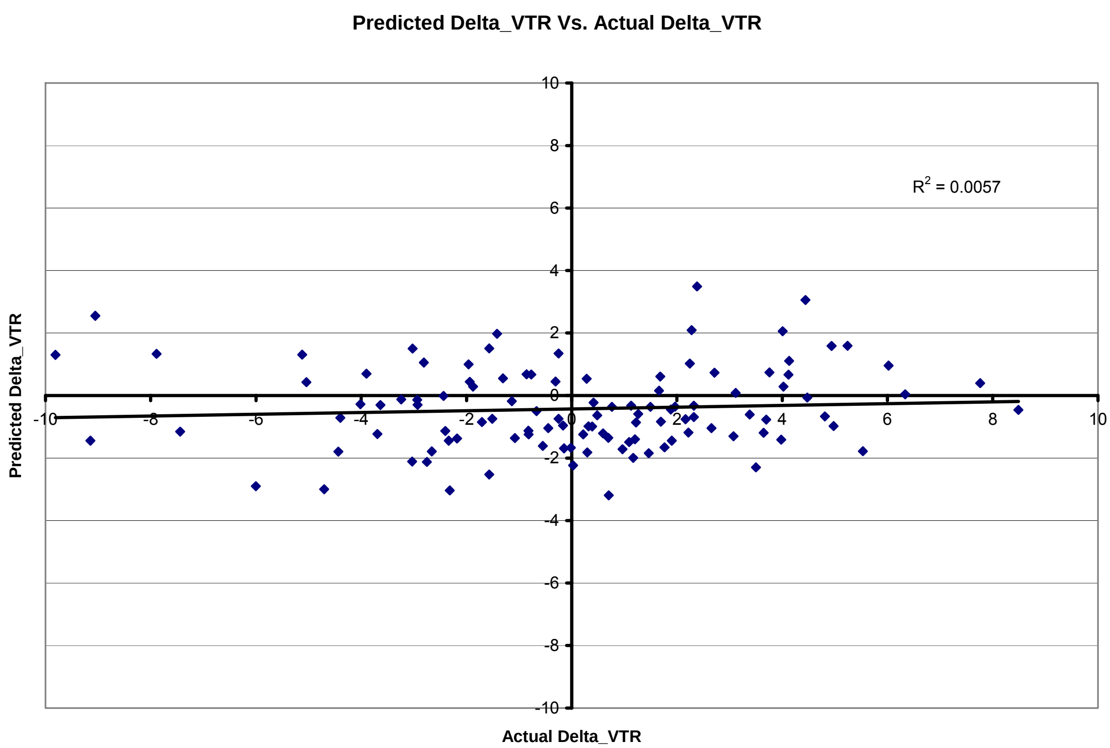


Figure 71: Scatter plot for Tucson validation set

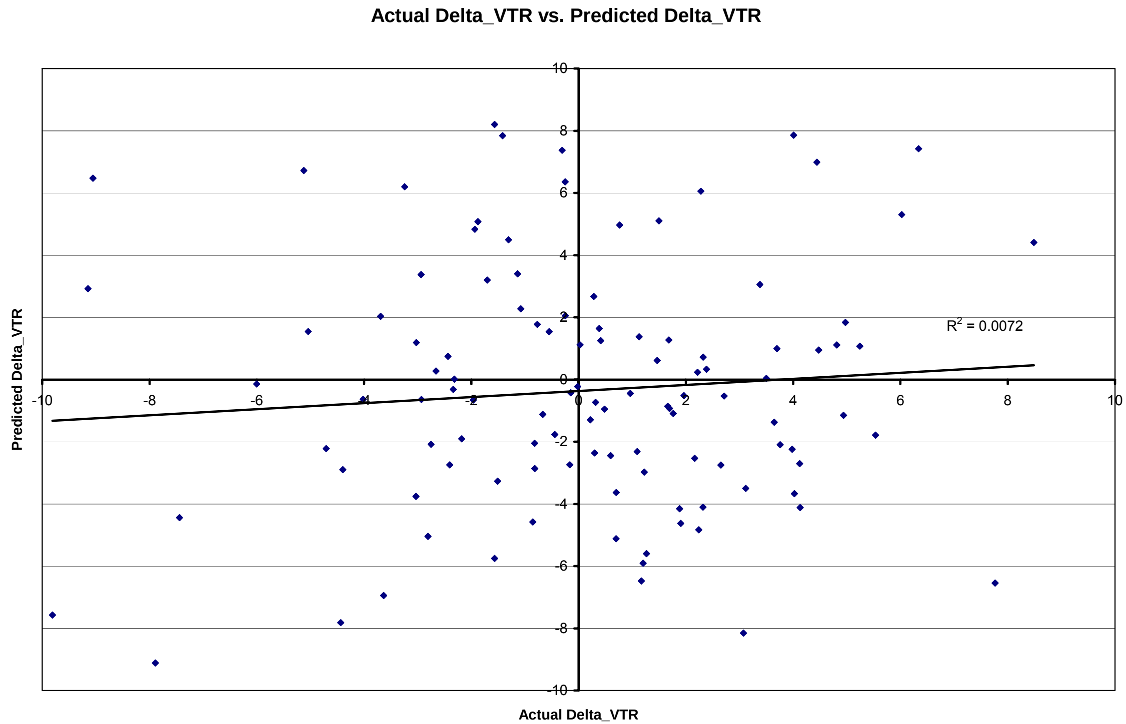


Figure 72: Scatter plot for neural network model built on over-sampled grouped Tucson data (recommended Tucson model)

Comparing the bin accuracy and R-square value charts, a conclusion was reached that equally sampled combined models were not able to improve the accuracy of the Los Angles and Washington data sets. The reason for such poor accuracy with the Los Angeles data may be due to such a small representation of Los Angeles data. Only 2,018 samples of the training data set were from Los Angeles. Inconsistency between Washington and the other two data sets could also be held responsible for the poor accuracy between Washington and the other data sets.

For the Tucson data, the accuracy in moderate range of change in VTR was improved with full range accuracy remaining constant at cost of reduction in the R-square as compared to the recommended model. But when examining the scatter plot, the neural net model that was built on equally sampled combined data only predicted small changes in VTR as compared to the recommended Tucson model, which predicted large changes in VTR. It was very difficult to decide between these two models and choose one as the best. Just like in the Washington data, when it was impossible to pick the best model, a cross-tab for positive/negative actual changes in VTR against positive/negative predicted changes in VTR on validation set was done.

Table 43 A-D: Validation Set

A

NN model on full sample ungrouped incentive dataset		
Actual Delta_VTR / Predicted Delta_VTR	**Negative**	**Positive**
Negative	25	31
Positive	27	29

B

NN model on equally sampled combined data		
Actual Delta_VTR / Predicted Delta_VTR	**Negative**	**Positive**
Negative	34	18
Positive	41	19

C

NN model on full sample ungrouped incentive dataset	
precision	0.446
recall	0.481
F-value	0.463

D

NN model on equally sampled combined data	
precision	**0.654**
recall	0.453
F-value	**0.535**

From the validation set cross-tabulations, it was decided that the neural network model that was built on equally sampled combined data was better than the independent model built on the over-sampled grouped incentives. The neural network model built on the equally sampled combined data resulted in better F-value on the validation set and therefore becomes the recommended model for Tucson.

V. WORKSITE TRIP REDUCTION MANUAL

MANUAL FOR COMBINED DATA MODEL (GENERAL MODEL)

The only model to get better results simultaneously on all three cities' validation sets was a neural network model built with no variable selection on equally sampled combined data. For this model, the training data from all of the three cities was combined in such a manner that an equal number (2,018) of records from each training data set contributed to the combined training data, thus eliminating the bias towards any dataset. The variables in this equally sampled combined dataset are shown in Table 46.

Table 44: Combined data variables and grouping

Variables	Description	Grouping
ALONESHARE	Alone Share	
TRANSITSHARE	Transit Share	
CVPOOLSHARE	Carpool +Vanpool Share	
WALKSHARE	Walk Share	
BCYCLESHARE	Bicycle Share	
MCYCLESHARE	Motorcycle Share	
TELESHARE	Telecommute Share	
CWW336	3/36 compressed work week share	
CWW440	4/40 compressed work week share	
CWW980	8/80 compressed work week share	
FACILITY_AMENITIES	facilities & amenities	Passenger Loading Areas Other Facility Improvements Preferential Parking Areas Bike Racks and Bike Lockers Shower and Lockers
GRH	Guaranteed ride home programs	TMA/TMO Provided Guaranteed Return Trip Company Vehicle Guaranteed Return Trip Emergencies Guaranteed Return Trip Other Guaranteed Return Trip Program Rental Car Guaranteed Return Trip Taxi Guaranteed Return Trip Unscheduled Overtime Guaranteed Return
FLEX	flexible timing	Flextime for Ride sharers (Work Shifts) Flextime for Ride sharers (Grace Period)
MRKT	Marketing programs	Commuter Information Center Commuter Fairs Focus Groups Posted Materials New Hire Orientation Other Marketing Elements Personal Communication Company Recognition Special Interest Club (Biking, Walking) TMA/TMO Membership Written Materials Zip Code Meetings
RS_MATCH	Ride share matching programs	Regional Commuter Management Agency Employer-Based Rideshare Matching System
FINANCIAL	financial incentives	Transportation Allowances On-Going Bike-to-Work Subsidies

		On-Going Carpooling Subsidies Other Direct Financial Subsidies On-Going Walk-to-Work Subsidies
PARKMGT	Parking management	Increased Parking Costs for Drive Alones Other Parking Management Strategies Subsidized Parking for Ride sharers
TELE	Telecommute program	Work at Home Work at Satellite Center
CWW	Compressed work week program	3/36, 4/40, 9/80 & other Compressed Work Week Schedule
ONSITE	onsite incentives	On-Site Childcare Service Other On-Site Services Cafeteria, ATM's, Postal, Fitness Center Transit Information or Pass Sales
DIRECT_NONFINAN	Non financial incentives	Auto Services (Fuel, Oil, Tune-Up) Gift Certificates Free Meals Other Direct Non-Financial Incentives Catalogue Points Additional Time Off with Pay Drawings, Free Meals, Certificates, etc
COMMTAX	commuter tax benefit incentives	Introductory Transit Passes or Subsidies Subsidized Vanpool Seats On-Going Transit Subsidies On-Going Vanpooling Subsidies
VTR	Vehicle trip rate	Vehicle trip rate
DELTA_VTR	Change in VTR (dependant variable)	Change in VTR (dependant variable)

For the purpose of creating a manual, the training data and the three city validation sets along with the unused training data from Los Angeles were combined into one file with 21,267 records. The neural network model with no variable selection built on equally sampled combined data was used to predict the change in Vehicle Trip Rate (VTR) for all of these records.

The Figures 73 and 74 show the number of records per Incentive plan combination and cumulative number of records versus distinct Incentive plans.

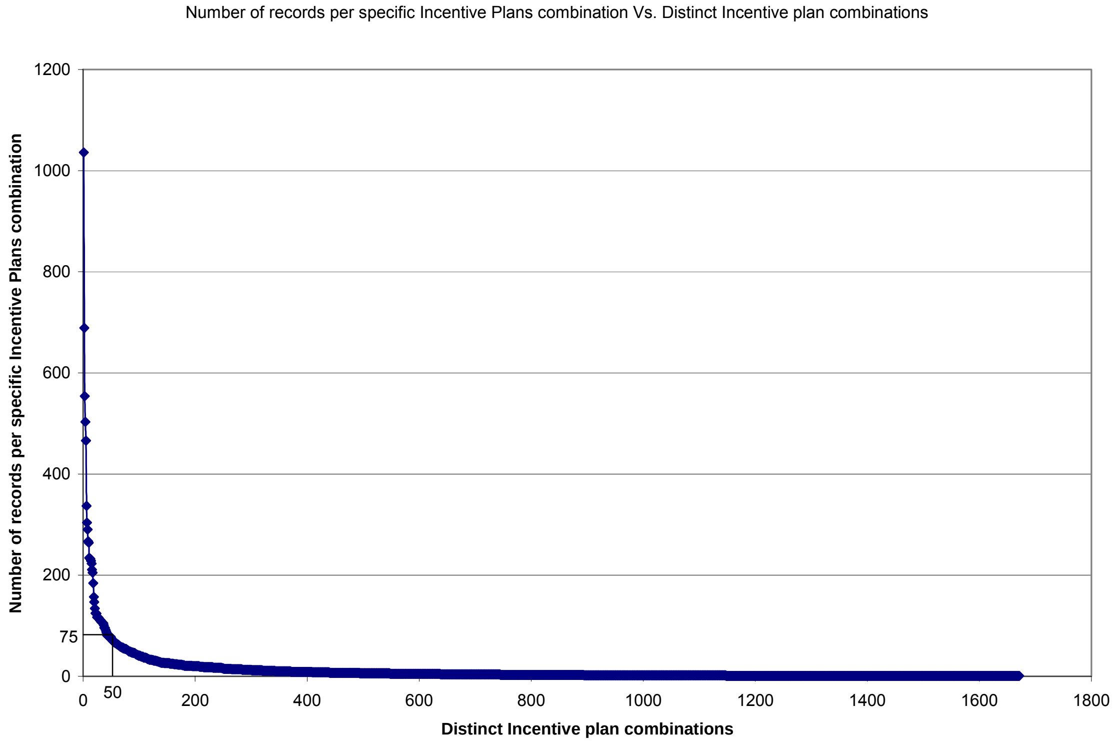


Figure 73: Number of records by incentive plan combinations

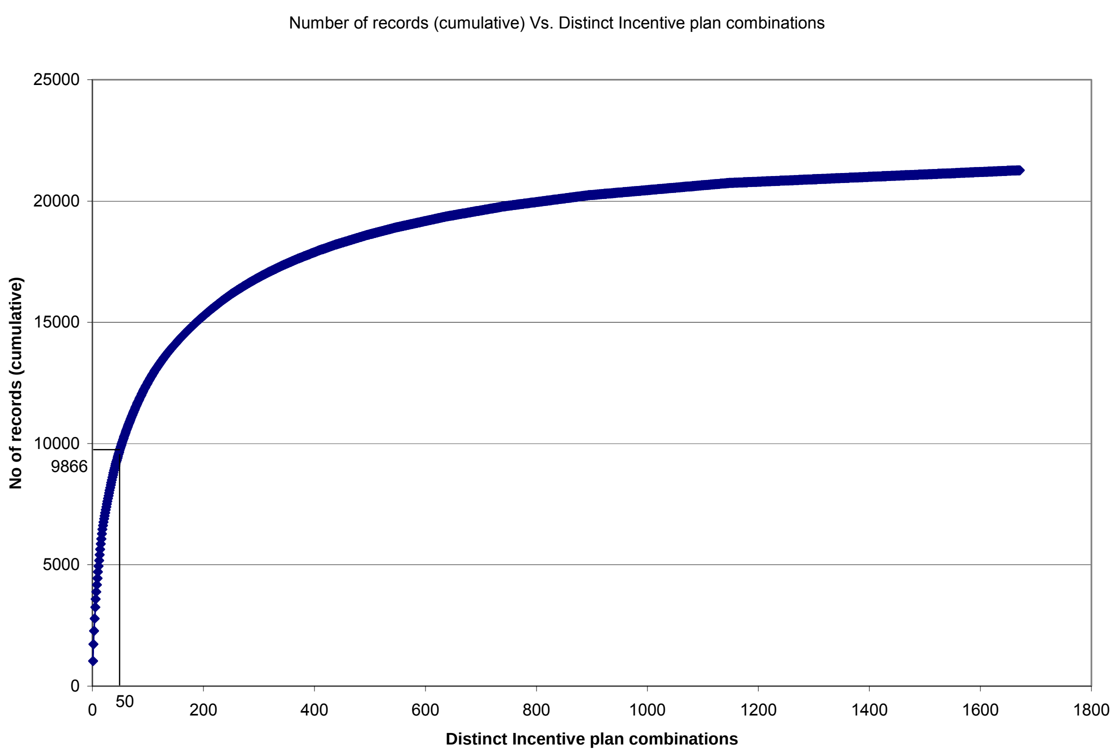


Figure 74: Number of records (cumulative) by plan combinations

It can be seen from Figures 74 and 75 that there are 1,671 distinct incentive plan combinations in total, and out of these, 50 combinations are implemented by at least 75 records. And these 50 distinct Incentive plan combinations have been implemented by 9,866 records in total. These are the incentive plan combinations that seem to be widely accepted and implemented by many worksites. Table 46 shows these 50 incentive plan combinations and the total number of records implementing each specific combination ("1" indicates plan offered).

Table 45: Widely implemented incentive plan combinations

Sr.No.	FACILITY_AMENITIES	onsite	commtax	FINANCIAL	CWW	FLEX	TELE	PARKMGT	direct_nonfinan	GRH	RS_MATCH	Mrkt	Total
1	√	√	√	√					√	√	√	√	1036
2	√	√	√	√		√			√	√	√	√	689
3	√	√	√						√	√	√	√	554
4	√		√	√					√	√	√	√	503
5	√	√							√	√	√	√	466
6		√	√	√					√	√	√	√	337
7	√	√				√			√	√	√	√	304
8	√	√	√	√	√				√	√	√	√	290
9	√		√						√	√	√	√	267
10	√	√			√				√	√	√	√	264
11	√		√	√						√	√	√	234
12	√									√	√	√	233
13	√	√	√			√			√	√	√	√	232
14	√	√	√	√						√	√	√	228
15	√								√	√	√	√	223
16	√	√	√	√	√	√			√	√	√	√	211
17	√	√	√	√					√	√		√	205
18	√	√	√		√				√	√	√	√	184
19	√		√	√		√			√	√	√	√	157
20			√	√					√	√	√	√	147
21	√	√			√		√		√	√	√	√	134
22	√		√			√			√	√	√	√	125
23	√										√	√	124
24	√	√		√					√	√	√	√	124
25	√	√	√		√	√			√	√	√	√	117
26	√	√	√	√		√				√	√	√	117
27	√	√									√	√	116
28	√	√	√		√		√		√	√	√	√	115
29	√	√	√						√	√		√	113
30	√		√	√	√				√	√	√	√	111
31	√	√								√	√	√	111
32	√					√			√	√	√	√	108
33		√	√						√	√	√	√	107
34	√	√		√	√	√			√	√	√	√	106
35	√	√	√							√	√	√	106
36	√		√	√					√	√		√	105
37	√			√					√	√	√	√	101
38	√	√	√	√	√		√		√	√	√	√	96
39	√	√	√	√	√	√	√		√	√	√	√	96
40	√			√						√		√	92
41	√	√							√		√	√	89
42	√	√			√		√			√	√	√	86
43	√	√			√		√		√		√	√	81
44	√	√	√	√			√		√	√	√	√	81
45	√	√			√						√	√	80
46	√				√	√				√	√	√	78
47	√		√	√	√	√			√	√	√	√	78
48	√	√					√				√	√	78
49	√	√	√	√				√	√	√	√	√	78
50			√						√	√	√	√	76

It can be seen from the table 46, that facilities & Amenities (showers & lockers, bike racks, etc.) and onsite incentives (onsite childcare, cafeteria, etc.) are dependant on infrastructure development were the worksite is located. So a transportation engineer using this manual can know it in advance if the new worksite under consideration has these incentives provided to their employees. Also from Transit infrastructure in the vicinity of the worksite, the baseline share of employees traveling by transit can be determined for the worksite.

So for example, if a new worksite expects its transit share to be in the range of 0% to 5% and has no knowledge about other shares, then assuming the other mode shares to be close to zero, the worksite might expect a vehicle trip rate in the range of 100% to 90%. Our manual provides tables with these kinds of scenarios, in this case table 49, containing our learned models averaged change in VTR predictions, the averaged change in VTR observed at the worksites, and the number of worksite records that matched the transit share & VTR scenario for the above stated incentive 50 incentive plan combinations. So for the above example, if the worksite offers all incentives except compressed work week, flexible timing, telecommuting and parking management incentives (Table 49, Sr.No.1), then according to our model the worksite can expect a reduction of 5.3 vehicle per 100 (i.e. -5.3 change in VTR), with 4.5 observed reduction over 238 records. The number of records can be used as a measure of confidence while using the predicted values. The all scenario tables in the manual use abbreviations for the names of the incentives as shown in table 48.

Table 46: Incentives codes

Incentives	Code
facilities & amenities	F&A
Guaranteed ride home programs	GRH
flexible timing	FLEX
Marketing programs	MRKT
Ride share matching programs	RSMP
financial incentives	FIN
Parking management	PMT
Telecommute program	TELE
Compressed work week program	CWW
onsite incentives	ONS
Non financial incentives	NONF
commuter tax benefit incentives	CTB

Procedure to look-up 'Change in VTR' prediction tables

1) Decide your Transit share range (Infrastructure dependant)
2) Decide your Vehicle Trip Rate range
3) Find the appropriate table
4) Look-up the table for incentive plan combinations and its predicted change in VTR (bold), actual observed change in VTR and number of records matching that criteria

Table 47: Model's change in VTR prediction for worksites with Transit share between 0% to 5% and Vehicle Trip Rate between 90 to 100

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0-0.05) VTR- [100-90)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	**-5.3**	-4.5	238
2	√	√	√	√		√			√	√	√	√	**-3.7**	-4.9	162
3	√	√	√						√	√	√	√	**-6.5**	-3.9	120
4	√		√	√					√	√	√	√	**-6.1**	-3.8	123
5	√	√							√	√	√	√	**-5.3**	-4.0	93
6		√	√	√					√	√	√	√	**-6.3**	-4.1	84
7	√	√				√			√	√	√	√	**-4.1**	-3.2	81
8	√	√	√	√	√				√	√	√	√	**-2.8**	-4.6	30
9	√		√						√	√	√	√	**-6.8**	-4.8	48
10	√	√			√				√	√	√	√	**-3.2**	-4.3	46
11	√		√	√						√	√	√	**-6.4**	-4.7	39
12	√									√	√	√	**-6.9**	-1.4	62
13	√	√	√			√			√	√	√	√	**-4.9**	-5.5	54
14	√	√	√	√						√	√	√	**-5.7**	-3.8	60
15	√								√	√	√	√	**-5.7**	-5.1	40
16	√	√	√	√	√	√			√	√	√	√	**-1.5**	-8.0	38
17	√	√	√	√					√	√		√	**-5.8**	-4.4	42
18	√	√	√		√				√	√	√	√	**-3.1**	-4.3	15
19	√		√	√		√			√	√	√	√	**-4.3**	-3.9	49
20			√	√					√	√	√	√	**-5.7**	-5.1	28
21	√	√			√		√		√	√	√	√	**-3.2**	-7.3	16
22	√		√			√			√	√	√	√	**-5.5**	-7.6	22
23	√										√	√	**-6.0**	-1.0	30
24	√	√		√					√	√	√	√	**-4.3**	-3.6	32
25	√	√	√		√	√			√	√	√	√	**-0.9**	-2.3	18
26	√	√	√	√		√				√	√	√	**-4.8**	-6.4	34
27	√	√									√	√	**-5.1**	-3.4	35
28	√	√	√		√		√		√	√	√	√	**-4.4**	-3.8	15
29	√	√	√						√	√		√	**-6.9**	-4.8	25
30	√		√	√	√				√	√	√	√	**-3.2**	-7.7	20
31	√	√								√	√	√	**-4.3**	-3.7	17
32	√					√			√	√	√	√	**-4.8**	-4.8	29
33		√	√						√	√	√	√	**-5.3**	-4.3	19
34	√	√		√	√	√			√	√	√	√	**-1.7**	-5.4	17
35	√	√	√							√	√	√	**-6.5**	-2.8	20
36	√		√	√					√	√		√	**-5.9**	-4.3	17
37	√			√					√	√	√	√	**-4.8**	-2.9	33
38	√	√	√	√	√		√		√	√	√	√	**-3.1**	-6.3	17
39	√	√	√	√	√	√	√		√	√	√	√	**-3.5**	-4.5	16
40	√			√						√		√	**-5.1**	-4.1	18
41	√	√							√		√	√	**-6.1**	-3.7	22
42	√	√			√		√			√	√	√	**-2.9**	-5.9	6
43	√	√			√		√		√		√	√	**-2.7**	-3.5	23
44	√	√	√	√			√		√	√	√	√	**-5.3**	-3.9	7
45	√	√			√						√	√	**-2.8**	-4.2	20
46	√				√	√				√	√	√	**-3.0**	-0.4	40
47	√		√	√	√	√			√	√	√	√	**-2.6**	-2.8	6
48	√	√					√				√	√	**-4.4**	-1.0	13
49	√	√	√	√				√	√	√	√	√	**-5.5**	-5.6	15
50			√						√	√	√	√	**-6.5**	-4.9	15

Table 48: Model's change in VTR prediction for worksites with Transit share between 0% to 5% and Vehicle Trip Rate between 80 to 90

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0-0.05) VTR- [90-80)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	**-2.9**	-1.4	461
2	√	√	√	√		√			√	√	√	√	**-1.6**	-1.9	342
3	√	√	√						√	√	√	√	**-3.7**	-1.2	192
4	√		√	√					√	√	√	√	**-4.2**	-1.9	215
5	√	√							√	√	√	√	**-2.1**	-0.9	174
6		√	√	√					√	√	√	√	**-3.2**	-1.7	115
7	√	√				√			√	√	√	√	**-1.9**	-2.3	126
8	√	√	√	√	√				√	√	√	√	**0.0**	-1.6	142
9	√		√						√	√	√	√	**-4.9**	-2.1	79
10	√	√			√				√	√	√	√	**-0.5**	-0.8	108
11	√		√	√						√	√	√	**-4.2**	-1.9	215
12	√									√	√	√	**-4.1**	-0.1	82
13	√	√	√			√			√	√	√	√	**-2.9**	-1.2	93
14	√	√	√	√						√	√	√	**-2.6**	-2.1	77
15	√								√	√	√	√	**-3.7**	-2.3	69
16	√	√	√	√	√	√			√	√	√	√	**0.6**	-1.8	103
17	√	√	√	√					√	√		√	**-3.3**	-0.9	89
18	√	√	√		√				√	√	√	√	**-0.9**	-2.6	75
19	√		√	√		√			√	√	√	√	**-3.4**	-3.3	64
20			√	√					√	√	√	√	**-3.6**	-1.7	49
21	√	√			√		√		√	√	√	√	**0.2**	-2.3	60
22	√		√			√			√	√	√	√	**-4.4**	-2.3	47
23	√										√	√	**-3.1**	1.3	45
24	√	√		√					√	√	√	√	**-1.4**	-0.5	59
25	√	√	√		√	√			√	√	√	√	**-0.2**	-0.4	52
26	√	√	√	√		√				√	√	√	**-1.3**	-3.0	47
27	√	√									√	√	**-2.7**	-1.4	21
28	√	√	√		√		√		√	√	√	√	**-0.6**	-1.2	51
29	√	√	√						√	√		√	**-3.7**	-2.5	32
30	√		√	√	√				√	√	√	√	**-1.6**	-1.6	47
31	√	√								√	√	√	**-2.0**	-3.0	22
32	√					√			√	√	√	√	**-3.3**	-2.4	32
33		√	√						√	√	√	√	**-4.1**	-3.2	40
34	√	√		√	√	√			√	√	√	√	**0.3**	0.2	60
35	√	√	√							√	√	√	**-2.9**	-1.7	35
36	√		√	√					√	√		√	**-3.5**	-2.7	52
37	√			√					√	√	√	√	**-2.9**	-2.1	42
38	√	√	√	√	√		√		√	√	√	√	**0.1**	-0.9	46
39	√	√	√	√	√	√	√		√	√	√	√	**-1.0**	-1.8	41
40	√			√						√		√	**-2.1**	-0.7	57
41	√	√							√		√	√	**-2.4**	-2.2	19
42	√	√			√		√			√	√	√	**-0.5**	-1.1	44
43	√	√			√		√		√		√	√	**0.7**	-1.5	36
44	√	√	√	√			√		√	√	√	√	**-2.1**	-0.4	45
45	√	√			√						√	√	**0.3**	-0.3	30
46	√				√	√				√	√	√	**-1.5**	0.8	19
47	√		√	√	√	√			√	√	√	√	**-1.6**	-3.9	34
48	√	√					√				√	√	**-3.1**	0.8	33
49	√	√	√	√				√	√	√	√	√	**-2.4**	-3.5	22
50			√						√	√	√	√	**-3.9**	-0.1	16

Table 49: Model's change in VTR prediction for worksites with Transit share between 0% to 5% and Vehicle Trip Rate between 70 to 80

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0-0.05) VTR- [80-70)		
													Predicted	**Actual**	**No.Ex**
1	√	√	√	√					√	√	√	√	**-0.6**	1.0	147
2	√	√	√	√		√			√	√	√	√	**0.4**	1.6	88
3	√	√	√						√	√	√	√	**-1.0**	0.5	93
4	√		√	√					√	√	√	√	**-1.6**	1.2	90
5	√	√							√	√	√	√	**-0.1**	-0.1	71
6		√	√	√					√	√	√	√	**-0.4**	-0.2	29
7	√	√				√			√	√	√	√	**0.8**	1.2	34
8	√	√	√	√	√				√	√	√	√	**1.6**	0.6	78
9	√		√						√	√	√	√	**-2.1**	1.2	44
10	√	√			√				√	√	√	√	**1.7**	-0.3	49
11	√		√	√						√	√	√	**-1.6**	1.2	90
12	√									√	√	√	**-1.0**	0.5	11
13	√	√	√			√			√	√	√	√	**-0.6**	1.2	28
14	√	√	√	√						√	√	√	**-0.1**	-0.5	26
15	√								√	√	√	√	**-1.1**	-0.9	49
16	√	√	√	√	√	√			√	√	√	√	**2.5**	0.4	41
17	√	√	√	√					√	√		√	**-2.1**	-3.6	23
18	√	√	√		√				√	√	√	√	**1.2**	2.2	44
19	√		√	√		√			√	√	√	√	**-1.6**	0.9	15
20			√	√					√	√	√	√	**-2.1**	1.8	16
21	√	√			√		√		√	√	√	√	**2.8**	-1.0	19
22	√		√			√			√	√	√	√	**-3.5**	-0.8	13
23	√										√	√	**-1.4**	-2.4	6
24	√	√		√					√	√	√	√	**-0.3**	0.8	18
25	√	√	√		√	√			√	√	√	√	**2.0**	0.9	25
26	√	√	√	√		√				√	√	√	**1.0**	0.2	11
27	√	√									√	√	**-0.4**	5.0	1
28	√	√	√		√		√		√	√	√	√	**1.8**	1.3	29
29	√	√	√						√	√		√	**-1.2**	3.6	18
30	√		√	√	√				√	√	√	√	**0.9**	-1.3	19
31	√	√								√	√	√	**0.0**	-2.2	9
32	√					√			√	√	√	√	**-1.2**	-0.7	17
33		√	√						√	√	√	√	**-0.9**	-2.9	10
34	√	√		√	√	√			√	√	√	√	**2.7**	0.7	13
35	√	√	√							√	√	√	**0.1**	0.3	10
36	√		√	√					√	√		√	**-1.4**	-1.5	15
37	√			√					√	√	√	√	**-0.5**	1.1	18
38	√	√	√	√	√		√		√	√	√	√	**2.4**	1.3	19
39	√	√	√	√	√	√	√		√	√	√	√	**2.1**	1.4	22
40	√			√						√		√	**-0.9**	-1.8	10
41	√	√							√		√	√	**1.0**	5.8	8
42	√	√			√		√			√	√	√	**2.5**	0.4	17
43	√	√			√		√		√		√	√	**2.9**	-2.4	8
44	√	√	√	√			√		√	√	√	√	**-0.6**	0.0	14
45	√	√			√						√	√	**0.8**	-2.8	13
46	√				√	√				√	√	√	**1.3**	1.4	12
47	√		√	√	√	√			√	√	√	√	**0.7**	0.7	23
48	√	√					√				√	√	**-0.9**	3.4	4
49	√	√	√	√				√	√	√	√	√	**0.7**	-2.4	8
50			√						√	√	√	√	**-1.9**	3.2	5

Table 50: Model's change in VTR prediction for worksites with Transit share between 0% to 5% and Vehicle Trip Rate between 60 to 70

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0-0.05) VTR- [70-60)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	**0.6**	2.5	27
2	√	√	√	√		√			√	√	√	√	**2.5**	2.0	16
3	√	√	√						√	√	√	√	**-0.1**	2.4	23
4	√		√	√					√	√	√	√	**1.3**	0.4	15
5	√	√							√	√	√	√	**1.5**	3.9	28
6		√	√	√					√	√	√	√	**-3.0**	-4.8	2
7	√	√				√			√	√	√	√	**2.7**	5.2	7
8	√	√	√	√	√				√	√	√	√	**4.5**	5.1	19
9	√		√						√	√	√	√	**0.2**	0.9	21
10	√	√			√				√	√	√	√	**3.9**	1.0	14
11	√		√	√						√	√	√	**1.3**	0.4	15
12	√									√	√	√	**1.5**	0.8	15
13	√	√	√			√			√	√	√	√	**1.0**	1.6	13
14	√	√	√	√						√	√	√	**1.5**	1.0	9
15	√								√	√	√	√	**1.1**	0.7	19
16	√	√	√	√	√	√			√	√	√	√	**4.3**	3.3	8
17	√	√	√	√					√	√		√	**-2.4**	3.8	3
18	√	√	√		√				√	√	√	√	**3.4**	0.8	20
19	√		√	√		√			√	√	√	√	**1.7**	2.1	9
20			√	√					√	√	√	√	**-1.0**	3.8	6
21	√	√			√		√		√	√	√	√	**5.0**	5.0	9
22	√		√			√			√	√	√	√	**-0.6**	3.4	5
23	√										√	√	**2.3**	0.5	10
24	√	√		√					√	√	√	√	**-0.4**	0.5	3
25	√	√	√		√	√			√	√	√	√	**2.8**	1.1	6
26	√	√	√	√		√				√	√	√	**0.9**	1.8	5
27	√	√									√	√	**1.5**	7.5	7
28	√	√	√		√		√		√	√	√	√	**3.1**	4.0	5
29	√	√	√						√	√		√	**-4.8**	7.8	1
30	√		√	√	√				√	√	√	√	**3.7**	-0.6	8
31	√	√								√	√	√	**1.2**	2.0	18
32	√					√			√	√	√	√	**0.5**	8.5	3
33		√	√						√	√	√	√	**0.5**	2.8	4
34	√	√		√	√	√			√	√	√	√	**-**	-	-
35	√	√	√							√	√	√	**0.3**	10.7	3
36	√		√	√					√	√		√	**0.1**	4.1	6
37	√			√					√	√	√	√	**2.3**	4.1	3
38	√	√	√	√	√		√		√	√	√	√	**4.9**	4.6	4
39	√	√	√	√	√	√	√		√	√	√	√	**1.0**	1.7	5
40	√			√						√		√	**0.9**	13.6	1
41	√	√							√		√	√	**3.2**	4.5	10
42	√	√			√		√			√	√	√	**4.1**	1.8	1
43	√	√			√		√		√		√	√	**-**	-	-
44	√	√	√	√			√		√	√	√	√	**1.5**	-0.8	2
45	√	√			√						√	√	**6.4**	10.9	1
46	√				√	√				√	√	√	**-**	-	-
47	√		√	√	√	√			√	√	√	√	**2.6**	4.7	7
48	√	√					√				√	√	**-2.4**	13.1	1
49	√	√	√	√				√	√	√	√	√	**2.0**	3.3	5
50			√						√	√	√	√	**1.0**	5.2	3

Table 51: Model's change in VTR prediction for worksites with Transit share between 0% to 5% and Vehicle Trip Rate between 50 to 60

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0-0.05) VTR- [60-50)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	**2.2**	5.5	5
2	√	√	√	√		√			√	√	√	√	**-1.9**	3.7	1
3	√	√	√						√	√	√	√	**3.1**	4.6	5
4	√		√	√					√	√	√	√	**-**	-	-
5	√	√							√	√	√	√	**3.5**	21.9	3
6		√	√	√					√	√	√	√	**-**	-	-
7	√	√				√			√	√	√	√	**2.1**	5.7	1
8	√	√	√	√	√				√	√	√	√	**6.6**	-0.2	2
9	√		√						√	√	√	√	**1.6**	5.2	2
10	√	√			√				√	√	√	√	**4.8**	1.4	1
11	√		√	√						√	√	√	**-**	-	-
12	√									√	√	√	**1.1**	-1.6	2
13	√	√	√			√			√	√	√	√	**7.6**	31.9	1
14	√	√	√	√						√	√	√	**-**	-	-
15	√								√	√	√	√	**3.1**	1.7	1
16	√	√	√	√	√	√			√	√	√	√	**5.9**	4.5	1
17	√	√	√	√					√	√		√	**2.6**	14.9	2
18	√	√	√		√				√	√	√	√	**2.4**	5.9	2
19	√		√	√		√			√	√	√	√	**-**	-	-
20			√	√					√	√	√	√	**-**	-	-
21	√	√			√		√		√	√	√	√	**3.5**	10.6	1
22	√		√			√			√	√	√	√	**-**	-	-
23	√										√	√	**4.0**	2.5	5
24	√	√		√					√	√	√	√	**-**	-	-
25	√	√	√		√	√			√	√	√	√	**-**	-	-
26	√	√	√	√		√				√	√	√	**-**	-	-
27	√	√									√	√	**2.9**	4.5	3
28	√	√	√		√		√		√	√	√	√	**8.8**	13.9	1
29	√	√	√						√	√		√	**3.0**	3.5	2
30	√		√	√	√				√	√	√	√	**-**	-	-
31	√	√								√	√	√	**4.2**	10.1	2
32	√					√			√	√	√	√	**-**	-	-
33		√	√						√	√	√	√	**-**	-	-
34	√	√		√	√	√			√	√	√	√	**-**	-	-
35	√	√	√							√	√	√	**-**	-	-
36	√		√	√					√	√		√	**1.3**	1.9	1
37	√			√					√	√	√	√	**-**	-	-
38	√	√	√	√	√		√		√	√	√	√	**5.3**	0.7	2
39	√	√	√	√	√	√	√		√	√	√	√	**-**	-	-
40	√			√						√		√	**2.8**	-5.3	1
41	√	√							√		√	√	**2.5**	4.6	2
42	√	√			√		√			√	√	√	**4.6**	-1.0	1
43	√	√			√		√		√		√	√	**-**	-	-
44	√	√	√	√			√		√	√	√	√	**-**	-	-
45	√	√			√						√	√	**-0.5**	21.3	1
46	√				√	√				√	√	√	**-**	-	-
47	√		√	√	√	√			√	√	√	√	**2.2**	10.4	2
48	√	√					√				√	√	**-**	-	-
49	√	√	√	√				√	√	√	√	√	**-**	-	-
50			√						√	√	√	√	**2.1**	2.8	1

Table 52: Model's change in VTR prediction for worksites with Transit share between 5% to 15% and Vehicle Trip Rate between 90 to 100

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.05-0.15) VTR- [100-90)		
													Predicted	**Actual**	**No.Ex**
1	√	√	√	√					√	√	√	√	**-6.8**	1.3	1
2	√	√	√	√		√			√	√	√	√	**-2.6**	-24.3	1
3	√	√	√						√	√	√	√	**-5.8**	2.3	1
4	√		√	√					√	√	√	√	**-3.3**	-5.4	1
5	√	√							√	√	√	√	**-4.6**	-1.3	2
6		√	√	√					√	√	√	√	**-7.5**	4.0	3
7	√	√				√			√	√	√	√	**-1.8**	-10.1	2
8	√	√	√	√	√				√	√	√	√	-	-	-
9	√		√						√	√	√	√	**-5.4**	-3.1	2
10	√	√			√				√	√	√	√	-	-	-
11	√		√	√						√	√	√	**-3.3**	-5.4	1
12	√									√	√	√	-	-	-
13	√	√	√			√			√	√	√	√	-	-	-
14	√	√	√	√						√	√	√	-	-	-
15	√								√	√	√	√	-	-	-
16	√	√	√	√	√	√			√	√	√	√	-	-	-
17	√	√	√	√					√	√		√	-	-	-
18	√	√	√		√				√	√	√	√	-	-	-
19	√		√	√		√			√	√	√	√	-	-	-
20			√	√					√	√	√	√	-	-	-
21	√	√			√		√		√	√	√	√	-	-	-
22	√		√			√			√	√	√	√	-	-	-
23	√										√	√	-	-	-
24	√	√		√					√	√	√	√	-	-	-
25	√	√	√		√	√			√	√	√	√	-	-	-
26	√	√	√	√		√				√	√	√	-	-	-
27	√	√									√	√	**0.5**	-5.0	3
28	√	√	√		√		√		√	√	√	√	-	-	-
29	√	√	√						√	√		√	-	-	-
30	√		√	√	√				√	√	√	√	-	-	-
31	√	√								√	√	√	-	-	-
32	√					√			√	√	√	√	-	-	-
33		√	√						√	√	√	√	-	-	-
34	√	√		√	√	√			√	√	√	√	-	-	-
35	√	√	√							√	√	√	-	-	-
36	√		√	√					√	√		√	**-7.6**	-0.9	1
37	√			√					√	√	√	√	-	-	-
38	√	√	√	√	√		√		√	√	√	√	-	-	-
39	√	√	√	√	√	√	√		√	√	√	√	-	-	-
40	√			√						√		√	-	-	-
41	√	√							√		√	√	**-2.9**	-9.9	1
42	√	√			√		√			√	√	√	-	-	-
43	√	√			√		√		√		√	√	-	-	-
44	√	√	√	√			√		√	√	√	√	-	-	-
45	√	√			√						√	√	-	-	-
46	√				√	√				√	√	√	-	-	-
47	√		√	√	√	√			√	√	√	√	-	-	-
48	√	√					√				√	√	-	-	-
49	√	√	√	√				√	√	√	√	√	-	-	-
50			√						√	√	√	√	**-4.9**	-11.0	1

Table 53: Model's change in VTR prediction for worksites with Transit share between 5% to 15% and Vehicle Trip Rate between 80 to 90

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.05-0.15) VTR- [90-80)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	**-3.2**	-3.7	44
2	√	√	√	√		√			√	√	√	√	**-1.1**	-1.1	27
3	√	√	√						√	√	√	√	**-4.5**	-2.6	26
4	√		√	√					√	√	√	√	**-4.3**	-5.7	14
5	√	√							√	√	√	√	**-1.9**	-3.7	20
6		√	√	√					√	√	√	√	**-3.7**	0.0	17
7	√	√				√			√	√	√	√	**-1.5**	-0.7	13
8	√	√	√	√	√				√	√	√	√	**0.5**	-17.2	3
9	√		√						√	√	√	√	**-4.5**	-4.8	10
10	√	√			√				√	√	√	√	**-1.2**	0.0	7
11	√		√	√						√	√	√	**-4.3**	-5.7	14
12	√									√	√	√	**-2.4**	0.2	8
13	√	√	√			√			√	√	√	√	**-2.3**	-3.7	14
14	√	√	√	√						√	√	√	**-2.7**	-2.6	15
15	√								√	√	√	√	**-3.4**	-0.1	12
16	√	√	√	√	√	√			√	√	√	√	**0.5**	-2.0	2
17	√	√	√	√					√	√		√	**-3.9**	0.5	13
18	√	√	√		√				√	√	√	√	**-1.8**	-3.8	5
19	√		√	√		√			√	√	√	√	**-1.6**	-4.7	6
20			√	√					√	√	√	√	**-4.4**	-3.0	13
21	√	√			√		√		√	√	√	√	**0.1**	-2.8	6
22	√		√			√			√	√	√	√	**-3.7**	-3.4	5
23	√										√	√	**-4.8**	-2.4	3
24	√	√		√					√	√	√	√	**-1.5**	-1.5	1
25	√	√	√		√	√			√	√	√	√	**0.0**	-0.5	6
26	√	√	√	√		√				√	√	√	**-3.2**	1.4	8
27	√	√									√	√	**-2.8**	-3.6	21
28	√	√	√		√		√		√	√	√	√	**-1.1**	-0.3	4
29	√	√	√						√	√		√	**-3.6**	-5.6	1
30	√		√	√	√				√	√	√	√	**-2.0**	0.0	7
31	√	√								√	√	√	**-1.2**	-3.6	3
32	√					√			√	√	√	√	**-1.6**	-0.4	5
33		√	√						√	√	√	√	**-3.7**	-1.1	10
34	√	√		√	√	√			√	√	√	√	**-0.4**	0.8	6
35	√	√	√							√	√	√	**-2.6**	-2.0	7
36	√		√	√					√	√		√	**-3.8**	-5.5	4
37	√			√					√	√	√	√	**-3.5**	-5.7	2
38	√	√	√	√	√		√		√	√	√	√	**0.8**	-4.9	3
39	√	√	√	√	√	√	√		√	√	√	√	**-0.9**	2.0	5
40	√			√						√		√	**-2.5**	-1.4	3
41	√	√							√		√	√	**-1.4**	4.6	3
42	√	√			√		√			√	√	√	**-2.9**	1.2	2
43	√	√			√		√		√		√	√	**0.9**	-3.8	6
44	√	√	√	√			√		√	√	√	√	**-3.7**	-7.4	4
45	√	√			√						√	√	**0.1**	-0.1	4
46	√				√	√				√	√	√	**-1.8**	-1.4	1
47	√		√	√	√	√			√	√	√	√	**-1.1**	-9.4	2
48	√	√					√				√	√	**-3.6**	0.5	4
49	√	√	√	√				√	√	√	√	√	**-3.5**	-3.0	3
50			√						√	√	√	√	**-4.4**	-0.2	7

Table 54: Model's change in VTR prediction for worksites with Transit share between 5% to 15% and Vehicle Trip Rate between 70 to 80

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.05-0.15) VTR- [80-70)		
													Predicted	**Actual**	**No.Ex**
1	√	√	√	√					√	√	√	√	**-1.5**	-0.2	62
2	√	√	√	√		√			√	√	√	√	**-0.2**	-1.5	24
3	√	√	√						√	√	√	√	**-1.4**	0.6	54
4	√		√	√					√	√	√	√	**-2.2**	-1.4	28
5	√	√							√	√	√	√	**0.1**	0.1	33
6		√	√	√					√	√	√	√	**-1.0**	0.5	52
7	√	√				√			√	√	√	√	**-0.7**	0.0	20
8	√	√	√	√	√				√	√	√	√	**1.1**	0.3	11
9	√		√						√	√	√	√	**-2.5**	0.2	30
10	√	√			√				√	√	√	√	**0.8**	0.6	29
11	√		√	√						√	√	√	**-2.2**	-1.4	28
12	√									√	√	√	**-0.2**	1.6	18
13	√	√	√			√			√	√	√	√	**-1.1**	-2.6	14
14	√	√	√	√						√	√	√	**-1.4**	-0.8	18
15	√								√	√	√	√	**-0.6**	-0.5	9
16	√	√	√	√	√	√			√	√	√	√	**1.4**	1.7	12
17	√	√	√	√					√	√		√	**-2.2**	2.1	16
18	√	√	√		√				√	√	√	√	**0.2**	0.6	15
19	√		√	√		√			√	√	√	√	**-1.8**	0.7	6
20			√	√					√	√	√	√	**-2.1**	0.3	18
21	√	√			√		√		√	√	√	√	**1.7**	-1.1	13
22	√		√			√			√	√	√	√	**-3.2**	-1.5	13
23	√										√	√	**-1.3**	-0.6	7
24	√	√		√					√	√	√	√	**-0.2**	-3.4	9
25	√	√	√		√	√			√	√	√	√	**1.1**	0.1	7
26	√	√	√	√		√				√	√	√	**0.8**	0.8	5
27	√	√									√	√	**-0.7**	-1.2	3
28	√	√	√		√		√		√	√	√	√	**0.2**	3.3	6
29	√	√	√						√	√		√	**-0.9**	0.8	14
30	√		√	√	√				√	√	√	√	**-0.2**	-3.5	5
31	√	√								√	√	√	**1.7**	2.9	2
32	√					√			√	√	√	√	**-1.4**	3.8	11
33		√	√						√	√	√	√	**0.1**	-3.0	9
34	√	√		√	√	√			√	√	√	√	**1.1**	-1.6	9
35	√	√	√							√	√	√	**-0.9**	1.1	12
36	√		√	√					√	√		√	**-0.3**	-5.5	3
37	√			√					√	√	√	√	**1.6**	-0.6	2
38	√	√	√	√	√		√		√	√	√	√	**3.4**	-1.5	4
39	√	√	√	√	√	√	√		√	√	√	√	**1.4**	0.4	5
40	√			√						√		√	**-**	-	-
41	√	√							√		√	√	**0.9**	-3.0	7
42	√	√			√		√			√	√	√	**2.0**	0.6	9
43	√	√			√		√		√		√	√	**2.5**	-6.0	4
44	√	√	√	√			√		√	√	√	√	**0.5**	0.3	7
45	√	√			√						√	√	**1.2**	-1.6	5
46	√				√	√				√	√	√	**-1.8**	0.9	4
47	√		√	√	√	√			√	√	√	√	**-0.4**	-4.2	2
48	√	√					√				√	√	**-2.8**	2.7	13
49	√	√	√	√				√	√	√	√	√	**-1.9**	-0.5	7
50			√						√	√	√	√	**-1.5**	-1.2	9

Table 55: Model's change in VTR prediction for worksites with Transit share between 5% to 15% and Vehicle Trip Rate between 60 to 70

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.05-0.15) VTR- [70-60)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	**-0.1**	2.0	16
2	√	√	√	√		√			√	√	√	√	**1.7**	4.2	8
3	√	√	√						√	√	√	√	**1.5**	-0.5	15
4	√		√	√					√	√	√	√	**0.3**	5.5	8
5	√	√							√	√	√	√	**1.9**	3.4	14
6		√	√	√					√	√	√	√	**0.5**	3.0	10
7	√	√				√			√	√	√	√	**2.9**	7.2	12
8	√	√	√	√	√				√	√	√	√	**-1.1**	5.7	4
9	√		√						√	√	√	√	**0.4**	2.1	11
10	√	√			√				√	√	√	√	**1.1**	0.7	7
11	√		√	√						√	√	√	**0.3**	5.5	8
12	√									√	√	√	**1.8**	1.9	16
13	√	√	√			√			√	√	√	√	**1.1**	1.8	9
14	√	√	√	√						√	√	√	**2.1**	5.7	6
15	√								√	√	√	√	**1.7**	2.1	10
16	√	√	√	√	√	√			√	√	√	√	**3.7**	-3.3	4
17	√	√	√	√					√	√		√	**-1.0**	2.0	8
18	√	√	√		√				√	√	√	√	**2.7**	5.3	4
19	√		√	√		√			√	√	√	√	**0.7**	4.1	5
20			√	√					√	√	√	√	**-1.3**	5.3	5
21	√	√			√		√		√	√	√	√	**2.1**	0.6	8
22	√		√			√			√	√	√	√	**-0.9**	4.7	11
23	√										√	√	**1.5**	2.2	7
24	√	√		√					√	√	√	√	**0.4**	-3.3	2
25	√	√	√		√	√			√	√	√	√	**3.1**	3.8	1
26	√	√	√	√		√				√	√	√	**2.6**	-6.6	2
27	√	√									√	√	**2.2**	2.4	4
28	√	√	√		√		√		√	√	√	√	**2.7**	2.9	2
29	√	√	√						√	√		√	**2.3**	0.0	5
30	√		√	√	√				√	√	√	√	**-1.3**	-3.4	1
31	√	√								√	√	√	**1.8**	-0.7	10
32	√					√			√	√	√	√	**1.9**	8.2	3
33		√	√						√	√	√	√	**0.6**	12.2	6
34	√	√		√	√	√			√	√	√	√	**2.5**	3.4	1
35	√	√	√							√	√	√	**1.6**	0.6	4
36	√		√	√					√	√		√	**-1.1**	1.5	2
37	√			√					√	√	√	√	**-**	-	-
38	√	√	√	√	√		√		√	√	√	√	**-**	-	-
39	√	√	√	√	√	√	√		√	√	√	√	**3.6**	13.1	1
40	√			√						√		√	**0.9**	11.4	1
41	√	√							√		√	√	**3.1**	3.8	8
42	√	√			√		√			√	√	√	**0.8**	-2.3	5
43	√	√			√		√		√		√	√	**-**	-	-
44	√	√	√	√			√		√	√	√	√	**-**	-	-
45	√	√			√						√	√	**0.0**	17.3	1
46	√				√	√				√	√	√	**1.1**	3.5	1
47	√		√	√	√	√			√	√	√	√	**-2.2**	-1.1	1
48	√	√					√				√	√	**-1.7**	13.1	5
49	√	√	√	√				√	√	√	√	√	**1.7**	-2.3	5
50			√						√	√	√	√	**0.3**	-0.4	10

Table 56: Model's change in VTR prediction for worksites with Transit share between 5% to 15% and Vehicle Trip Rate between 50 to 60

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.05-0.15) VTR- [60-50)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	**3.7**	1.4	1
2	√	√	√	√		√			√	√	√	√	**0.4**	0.7	1
3	√	√	√						√	√	√	√	**2.4**	5.3	2
4	√		√	√					√	√	√	√	**-**	-	-
5	√	√							√	√	√	√	**0.7**	6.8	3
6		√	√	√					√	√	√	√	**-**	-	-
7	√	√				√			√	√	√	√	**6.3**	7.4	1
8	√	√	√	√	√				√	√	√	√	**-**	-	-
9	√		√						√	√	√	√	**-1.0**	2.3	3
10	√	√			√				√	√	√	√	**6.0**	12.0	1
11	√		√	√						√	√	√	**-**	-	-
12	√									√	√	√	**4.1**	5.0	4
13	√	√	√			√			√	√	√	√	**-**	-	-
14	√	√	√	√						√	√	√	**1.3**	2.0	3
15	√								√	√	√	√	**5.2**	4.7	2
16	√	√	√	√	√	√			√	√	√	√	**-**	-	-
17	√	√	√	√					√	√		√	**-**	-	-
18	√	√	√		√				√	√	√	√	**4.4**	9.9	2
19	√		√	√		√			√	√	√	√	**-**	-	-
20			√	√					√	√	√	√	**1.3**	6.9	1
21	√	√			√		√		√	√	√	√	**-**	-	-
22	√		√			√			√	√	√	√	**1.0**	7.1	4
23	√										√	√	**4.5**	2.0	3
24	√	√		√					√	√	√	√	**-**	-	-
25	√	√	√		√	√			√	√	√	√	**-**	-	-
26	√	√	√	√		√				√	√	√	**1.1**	5.0	1
27	√	√									√	√	**-**	-	-
28	√	√	√		√		√		√	√	√	√	**-**	-	-
29	√	√	√						√	√		√	**0.4**	6.4	3
30	√		√	√	√				√	√	√	√	**3.0**	2.1	2
31	√	√								√	√	√	**4.3**	-1.8	3
32	√					√			√	√	√	√	**3.3**	5.1	3
33		√	√						√	√	√	√	**4.1**	-0.6	3
34	√	√		√	√	√			√	√	√	√	**-**	-	-
35	√	√	√							√	√	√	**4.8**	4.6	4
36	√		√	√					√	√		√	**0.5**	-5.2	2
37	√			√					√	√	√	√	**3.8**	12.9	1
38	√	√	√	√	√		√		√	√	√	√	**-**	-	-
39	√	√	√	√	√	√	√		√	√	√	√	**-**	-	-
40	√			√						√		√	**-**	-	-
41	√	√							√		√	√	**4.6**	4.2	2
42	√	√			√		√			√	√	√	**-**	-	-
43	√	√			√		√		√		√	√	**-**	-	-
44	√	√	√	√			√		√	√	√	√	**-**	-	-
45	√	√			√						√	√	**-**	-	-
46	√				√	√				√	√	√	**-**	-	-
47	√		√	√	√	√			√	√	√	√	**-**	-	-
48	√	√					√				√	√	**-1.6**	17.3	2
49	√	√	√	√				√	√	√	√	√	**5.2**	2.7	1
50			√						√	√	√	√	**1.9**	-2.7	1

Table 57: Model's change in VTR prediction for worksites with Transit share between 15% to 25% and Vehicle Trip Rate between 70 to 80

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.15-0.25) VTR- [80-70)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	-3.2	-5.0	5
2	√	√	√	√		√			√	√	√	√	-1.9	-4.4	4
3	√	√	√						√	√	√	√	-3.5	-4.5	7
4	√		√	√					√	√	√	√	-3.3	6.1	2
5	√	√							√	√	√	√	-1.6	1.1	5
6		√	√	√					√	√	√	√	-3.5	0.7	8
7	√	√				√			√	√	√	√	-2.4	-10.7	1
8	√	√	√	√	√				√	√	√	√	-	-	-
9	√		√						√	√	√	√	-6.3	-6.2	1
10	√	√			√				√	√	√	√	-1.0	2.2	1
11	√		√	√						√	√	√	-3.3	6.1	2
12	√									√	√	√	-	-	-
13	√	√	√			√			√	√	√	√	-2.0	-5.5	1
14	√	√	√	√						√	√	√	-5.6	4.9	1
15	√								√	√	√	√	-1.7	-2.2	1
16	√	√	√	√	√	√			√	√	√	√	-	-	-
17	√	√	√	√					√	√		√	-2.7	-1.9	3
18	√	√	√		√				√	√	√	√	-	-	-
19	√		√	√		√			√	√	√	√	-	-	-
20			√	√					√	√	√	√	-3.7	-0.8	3
21	√	√			√		√		√	√	√	√	-	-	-
22	√		√			√			√	√	√	√	-	-	-
23	√										√	√	-	-	-
24	√	√		√					√	√	√	√	-	-	-
25	√	√	√		√	√			√	√	√	√	-0.7	6.4	1
26	√	√	√	√		√				√	√	√	-1.6	9.5	1
27	√	√									√	√	0.7	4.4	2
28	√	√	√		√		√		√	√	√	√	-	-	-
29	√	√	√						√	√		√	-0.8	-8.7	2
30	√		√	√	√				√	√	√	√	-0.9	4.7	1
31	√	√								√	√	√	-1.1	-3.7	1
32	√					√			√	√	√	√	-3.6	-11.5	3
33		√	√						√	√	√	√	-5.1	-3.0	2
34	√	√		√	√	√			√	√	√	√	-	-	-
35	√	√	√							√	√	√	-1.0	2.0	1
36	√		√	√					√	√		√	-	-	-
37	√			√					√	√	√	√	-	-	-
38	√	√	√	√	√		√		√	√	√	√	-	-	-
39	√	√	√	√	√	√	√		√	√	√	√	-	-	-
40	√			√						√		√	-	-	-
41	√	√							√		√	√	3.0	-3.0	1
42	√	√			√		√			√	√	√	-	-	-
43	√	√			√		√		√		√	√	2.5	5.9	2
44	√	√	√	√			√		√	√	√	√	-	-	-
45	√	√			√						√	√	-	-	-
46	√				√	√				√	√	√	-	-	-
47	√		√	√	√	√			√	√	√	√	-3.5	-1.8	1
48	√	√					√				√	√	-	-	-
49	√	√	√	√				√	√	√	√	√	-1.2	5.7	1
50			√						√	√	√	√	-1.1	-1.2	1

Table 58: Model's change in VTR prediction for worksites with Transit share between 15% to 25% and Vehicle Trip Rate between 60 to 70

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.15-0.25) VTR- [70-60)		
													Predicted	**Actual**	**No.Ex**
1	√	√	√	√					√	√	√	√	**-0.7**	1.6	12
2	√	√	√	√		√			√	√	√	√	**-0.1**	5.6	9
3	√	√	√						√	√	√	√	**-0.4**	1.2	7
4	√		√	√					√	√	√	√	**-2.0**	-3.7	4
5	√	√							√	√	√	√	**2.0**	0.4	13
6		√	√	√					√	√	√	√	**-1.1**	5.4	9
7	√	√				√			√	√	√	√	**1.0**	1.3	3
8	√	√	√	√	√				√	√	√	√	**-**	-	-
9	√		√						√	√	√	√	**-0.8**	0.2	7
10	√	√			√				√	√	√	√	**2.0**	-2.8	1
11	√		√	√						√	√	√	**-2.0**	-3.7	4
12	√									√	√	√	**1.3**	2.1	7
13	√	√	√			√			√	√	√	√	**-0.2**	0.4	2
14	√	√	√	√						√	√	√	**-1.0**	3.2	6
15	√								√	√	√	√	**0.4**	-11.8	2
16	√	√	√	√	√	√			√	√	√	√	**0.4**	-4.5	1
17	√	√	√	√					√	√		√	**-1.4**	1.5	5
18	√	√	√		√				√	√	√	√	**3.3**	-4.1	1
19	√		√	√		√			√	√	√	√	**-1.0**	8.2	2
20			√	√					√	√	√	√	**0.5**	-0.3	6
21	√	√			√		√		√	√	√	√	**4.8**	14.7	1
22	√		√			√			√	√	√	√	**-2.5**	0.2	4
23	√										√	√	**1.3**	8.2	1
24	√	√		√					√	√	√	√	**-**	-	-
25	√	√	√		√	√			√	√	√	√	**-**	-	-
26	√	√	√	√		√				√	√	√	**-0.1**	0.4	1
27	√	√									√	√	**-**	-	-
28	√	√	√		√		√		√	√	√	√	**-0.3**	-13.7	1
29	√	√	√						√	√		√	**-0.2**	2.4	7
30	√		√	√	√				√	√	√	√	**-0.3**	1.0	1
31	√	√								√	√	√	**0.4**	0.0	6
32	√					√			√	√	√	√	**-0.9**	8.8	1
33		√	√						√	√	√	√	**0.1**	8.1	2
34	√	√		√	√	√			√	√	√	√	**-**	-	-
35	√	√	√							√	√	√	**-0.8**	2.7	6
36	√		√	√					√	√		√	**-1.6**	-1.9	1
37	√			√					√	√	√	√	**-**	-	-
38	√	√	√	√	√		√		√	√	√	√	**-**	-	-
39	√	√	√	√	√	√	√		√	√	√	√	**-**	-	-
40	√			√						√		√	**-**	-	-
41	√	√							√		√	√	**2.1**	-5.3	2
42	√	√			√		√			√	√	√	**-**	-	-
43	√	√			√		√		√		√	√	**2.7**	-6.3	2
44	√	√	√	√			√		√	√	√	√	**0.0**	4.7	1
45	√	√			√						√	√	**-**	-	-
46	√				√	√				√	√	√	**-**	-	-
47	√		√	√	√	√			√	√	√	√	**-**	-	-
48	√	√					√				√	√	**-3.7**	-31.7	1
49	√	√	√	√				√	√	√	√	√	**-1.7**	-1.8	5
50			√						√	√	√	√	**-0.4**	1.1	4

Table 59: Model's change in VTR prediction for worksites with Transit share between 15% to 25% and Vehicle Trip Rate between 50 to 60

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.15-0.25) VTR- [60-50)		
													Predicted	**Actual**	**No.Ex**
1	√	√	√	√					√	√	√	√	**2.4**	26.4	2
2	√	√	√	√		√			√	√	√	√	**0.9**	6.8	1
3	√	√	√						√	√	√	√	**-0.3**	-4.3	1
4	√		√	√					√	√	√	√	-	-	-
5	√	√							√	√	√	√	**2.3**	-1.5	4
6		√	√	√					√	√	√	√	-	-	-
7	√	√				√			√	√	√	√	**4.4**	11.5	2
8	√	√	√	√	√				√	√	√	√	-	-	-
9	√		√						√	√	√	√	**0.2**	5.2	2
10	√	√			√				√	√	√	√	-	-	-
11	√		√	√						√	√	√	-	-	-
12	√									√	√	√	**3.3**	9.9	2
13	√	√	√			√			√	√	√	√	-	-	-
14	√	√	√	√						√	√	√	-	-	-
15	√								√	√	√	√	**4.2**	0.3	3
16	√	√	√	√	√	√			√	√	√	√	-	-	-
17	√	√	√	√					√	√		√	-	-	-
18	√	√	√		√				√	√	√	√	-	-	-
19	√		√	√		√			√	√	√	√	-	-	-
20			√	√					√	√	√	√	-	-	-
21	√	√			√		√		√	√	√	√	-	-	-
22	√		√			√			√	√	√	√	**2.4**	14.9	1
23	√										√	√	-	-	-
24	√	√		√					√	√	√	√	-	-	-
25	√	√	√		√	√			√	√	√	√	-	-	-
26	√	√	√	√		√				√	√	√	-	-	-
27	√	√									√	√	**1.0**	0.0	1
28	√	√	√		√		√		√	√	√	√	-	-	-
29	√	√	√						√	√		√	-	-	-
30	√		√	√	√				√	√	√	√	-	-	-
31	√	√								√	√	√	**3.2**	-3.4	4
32	√					√			√	√	√	√	-	-	-
33		√	√						√	√	√	√	**-2.0**	-12.5	1
34	√	√		√	√	√			√	√	√	√	-	-	-
35	√	√	√							√	√	√	**3.8**	4.8	1
36	√		√	√					√	√		√	-	-	-
37	√			√					√	√	√	√	-	-	-
38	√	√	√	√	√		√		√	√	√	√	-	-	-
39	√	√	√	√	√	√	√		√	√	√	√	-	-	-
40	√			√						√		√	-	-	-
41	√	√							√		√	√	**3.2**	11.7	1
42	√	√			√		√			√	√	√	-	-	-
43	√	√			√		√		√		√	√	-	-	-
44	√	√	√	√			√		√	√	√	√	-	-	-
45	√	√			√						√	√	**1.1**	-5.4	1
46	√				√	√				√	√	√	**8.0**	21.1	1
47	√		√	√	√	√			√	√	√	√	-	-	-
48	√	√					√				√	√	**-4.6**	22.1	1
49	√	√	√	√				√	√	√	√	√	**0.3**	-2.6	3
50			√						√	√	√	√	-	-	-

Table 60: Model's change in VTR prediction for worksites with Transit share between 25% to 35% and Vehicle Trip Rate between 60 to 70

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.25-0.35) VTR- [70-60)		
													Predicted	**Actual**	**No.Ex**
1	√	√	√	√					√	√	√	√	**-1.9**	7.2	3
2	√	√	√	√		√			√	√	√	√	**-1.9**	-3.9	4
3	√	√	√						√	√	√	√	**-3.0**	-7.3	1
4	√		√	√					√	√	√	√	**-1.9**	12.8	1
5	√	√							√	√	√	√	**-**	-	-
6		√	√	√					√	√	√	√	**-2.9**	-4.4	3
7	√	√				√			√	√	√	√	**-**	-	-
8	√	√	√	√	√				√	√	√	√	**-**	-	-
9	√		√						√	√	√	√	**-1.8**	-0.4	3
10	√	√			√				√	√	√	√	**-**	-	-
11	√		√	√						√	√	√	**-1.9**	12.8	1
12	√									√	√	√	**-3.3**	-8.2	2
13	√	√	√			√			√	√	√	√	**-3.1**	1.5	1
14	√	√	√	√						√	√	√	**-2.5**	-3.0	2
15	√								√	√	√	√	**-**	-	-
16	√	√	√	√	√	√			√	√	√	√	**-**	-	-
17	√	√	√	√					√	√		√	**-**	-	-
18	√	√	√		√				√	√	√	√	**-1.9**	1.5	1
19	√		√	√		√			√	√	√	√	**-**	-	-
20			√	√					√	√	√	√	**-**	-	-
21	√	√			√		√		√	√	√	√	**-**	-	-
22	√		√			√			√	√	√	√	**-**	-	-
23	√										√	√	**-**	-	-
24	√	√		√					√	√	√	√	**-**	-	-
25	√	√	√		√	√			√	√	√	√	**-**	-	-
26	√	√	√	√		√				√	√	√	**-3.2**	-8.5	1
27	√	√									√	√	**-**	-	-
28	√	√	√		√		√		√	√	√	√	**-**	-	-
29	√	√	√						√	√		√	**-2.4**	4.5	1
30	√		√	√	√				√	√	√	√	**-**	-	-
31	√	√								√	√	√	**-1.4**	-6.0	1
32	√					√			√	√	√	√	**-2.6**	6.7	1
33		√	√						√	√	√	√	**-**	-	-
34	√	√		√	√	√			√	√	√	√	**-**	-	-
35	√	√	√							√	√	√	**-1.7**	-7.3	2
36	√		√	√					√	√		√	**-**	-	-
37	√			√					√	√	√	√	**-**	-	-
38	√	√	√	√	√		√		√	√	√	√	**-**	-	-
39	√	√	√	√	√	√	√		√	√	√	√	**2.1**	-6.7	1
40	√			√						√		√	**-**	-	-
41	√	√							√		√	√	**-**	-	-
42	√	√			√		√			√	√	√	**-**	-	-
43	√	√			√		√		√		√	√	**-**	-	-
44	√	√	√	√			√		√	√	√	√	**-**	-	-
45	√	√			√						√	√	**-**	-	-
46	√				√	√				√	√	√	**-**	-	-
47	√		√	√	√	√			√	√	√	√	**-**	-	-
48	√	√					√				√	√	**-**	-	-
49	√	√	√	√				√	√	√	√	√	**-1.5**	-0.1	1
50			√						√	√	√	√	**-**	-	-

Table 61: Model's change in VTR prediction for worksites with Transit share between 25% to 35% and Vehicle Trip Rate between 50 to 60

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.25-0.35) VTR- [60-50)		
													Predicted	**Actual**	**No.Ex**
1	√	√	√	√					√	√	√	√	**-0.5**	-0.3	5
2	√	√	√	√		√			√	√	√	√	**-**	-	-
3	√	√	√						√	√	√	√	**0.6**	4.9	3
4	√		√	√					√	√	√	√	**-**	-	-
5	√	√							√	√	√	√	**-**	-	-
6		√	√	√					√	√	√	√	**0.1**	1.5	3
7	√	√				√			√	√	√	√	**-1.9**	6.8	1
8	√	√	√	√	√				√	√	√	√	**-**	-	-
9	√		√						√	√	√	√	**-1.4**	0.0	2
10	√	√			√				√	√	√	√	**-**	-	-
11	√		√	√						√	√	√	**-**	-	-
12	√									√	√	√	**-**	-	-
13	√	√	√			√			√	√	√	√	**1.3**	2.7	1
14	√	√	√	√						√	√	√	**1.5**	0.1	1
15	√								√	√	√	√	**1.5**	0.6	3
16	√	√	√	√	√	√			√	√	√	√	**-**	-	-
17	√	√	√	√					√	√		√	**-**	-	-
18	√	√	√		√				√	√	√	√	**-**	-	-
19	√		√	√		√			√	√	√	√	**-0.5**	4.2	1
20			√	√					√	√	√	√	**0.8**	-2.4	1
21	√	√			√		√		√	√	√	√	**5.3**	10.8	1
22	√		√			√			√	√	√	√	**-**	-	-
23	√										√	√	**-**	-	-
24	√	√		√					√	√	√	√	**-**	-	-
25	√	√	√		√	√			√	√	√	√	**-1.8**	29.3	1
26	√	√	√	√		√				√	√	√	**-**	-	-
27	√	√									√	√	**3.2**	0.1	3
28	√	√	√		√		√		√	√	√	√	**-**	-	-
29	√	√	√						√	√		√	**3.4**	8.8	1
30	√		√	√	√				√	√	√	√	**-**	-	-
31	√	√								√	√	√	**2.3**	3.2	2
32	√					√			√	√	√	√	**-**	-	-
33		√	√						√	√	√	√	**-2.0**	0.9	1
34	√	√		√	√	√			√	√	√	√	**-**	-	-
35	√	√	√							√	√	√	**2.1**	8.4	1
36	√		√	√					√	√		√	**-**	-	-
37	√			√					√	√	√	√	**-**	-	-
38	√	√	√	√	√		√		√	√	√	√	**-**	-	-
39	√	√	√	√	√	√	√		√	√	√	√	**-**	-	-
40	√			√						√		√	**1.1**	2.1	1
41	√	√							√		√	√	**-**	-	-
42	√	√			√		√			√	√	√	**-**	-	-
43	√	√			√		√		√		√	√	**-**	-	-
44	√	√	√	√			√		√	√	√	√	**0.3**	-5.8	1
45	√	√			√						√	√	**-**	-	-
46	√				√	√				√	√	√	**-**	-	-
47	√		√	√	√	√			√	√	√	√	**-**	-	-
48	√	√					√				√	√	**-**	-	-
49	√	√	√	√				√	√	√	√	√	**1.0**	-2.6	2
50			√						√	√	√	√	**2.0**	13.0	1

Table 62: Model’s change in VTR prediction for worksites with Transit share between 25% to 35% and Vehicle Trip Rate between 50 to 60

Sr. No.	F&A	ONS	CTB	FIN	CWW	FLEX	TELE	PMT	NONF	GRH	RSMP	MRKT	Transit Share [0.35-0.45) VTR- [60-50)		
													Predicted	Actual	No.Ex
1	√	√	√	√					√	√	√	√	**-1.8**	1.8	3
2	√	√	√	√		√			√	√	√	√	**-**	-	-
3	√	√	√						√	√	√	√	**-2.1**	-2.0	1
4	√		√	√					√	√	√	√	**-**	-	-
5	√	√							√	√	√	√	**-**	-	-
6		√	√	√					√	√	√	√	**-**	-	-
7	√	√				√			√	√	√	√	**-**	-	-
8	√	√	√	√	√				√	√	√	√	**-**	-	-
9	√		√						√	√	√	√	**-**	-	-
10	√	√			√				√	√	√	√	**-**	-	-
11	√		√	√						√	√	√	**-**	-	-
12	√									√	√	√	**-**	-	-
13	√	√	√			√			√	√	√	√	**-**	-	-
14	√	√	√	√						√	√	√	**-**	-	-
15	√								√	√	√	√	**1.4**	3.4	1
16	√	√	√	√	√	√			√	√	√	√	**-**	-	-
17	√	√	√	√					√	√		√	**-**	-	-
18	√	√	√		√				√	√	√	√	**-**	-	-
19	√		√	√		√			√	√	√	√	**-**	-	-
20			√	√					√	√	√	√	**-**	-	-
21	√	√			√		√		√	√	√	√	**-**	-	-
22	√		√			√			√	√	√	√	**-**	-	-
23	√										√	√	**-**	-	-
24	√	√		√					√	√	√	√	**-**	-	-
25	√	√	√		√	√			√	√	√	√	**-**	-	-
26	√	√	√	√		√				√	√	√	**-**	-	-
27	√	√									√	√	**-**	-	-
28	√	√	√		√		√		√	√	√	√	**-**	-	-
29	√	√	√						√	√		√	**-1.6**	8.5	1
30	√		√	√	√				√	√	√	√	**-**	-	-
31	√	√								√	√	√	**-0.9**	8.3	1
32	√					√			√	√	√	√	**-**	-	-
33		√	√						√	√	√	√	**-**	-	-
34	√	√		√	√	√			√	√	√	√	**-**	-	-
35	√	√	√							√	√	√	**-**	-	-
36	√		√	√					√	√		√	**-**	-	-
37	√			√					√	√	√	√	**-**	-	-
38	√	√	√	√	√		√		√	√	√	√	**-**	-	-
39	√	√	√	√	√	√	√		√	√	√	√	**-**	-	-
40	√			√						√		√	**-**	-	-
41	√	√							√		√	√	**-**	-	-
42	√	√			√		√			√	√	√	**-**	-	-
43	√	√			√		√		√		√	√	**-**	-	-
44	√	√	√	√			√		√	√	√	√	**-**	-	-
45	√	√			√						√	√	**-**	-	-
46	√				√	√				√	√	√	**-**	-	-
47	√		√	√	√	√			√	√	√	√	**-**	-	-
48	√	√					√				√	√	**-**	-	-
49	√	√	√	√				√	√	√	√	√	**-**	-	-
50			√						√	√	√	√	**-**	-	-

VI. SUMMARY

This project was undertaken with purpose of simplifying the transportation management process for all parties in the development and planning process for both residential and commercial enterprises. The software model developed from this project provides a valuable and interactive tool for anyone involved in the decision-making process for new developments and transportation programs. This software allows for precise prediction to the extent each incentive, disincentive, or program has an effect on vehicle trip rate.

This software model takes much of the guesswork out of planning developmental impact and roadway level of service. The user of the software can enter data for several variables, such as the number of employees, the shares of different commuting modes used by the employees, incentives offered, disincentives charged, etc. The learned model built on the data obtained from various sources culled from years of longitudinal research, returns a predictive answer as to the amount the vehicle trips that will be reduced or increased.

This project used several thousand worksite trip reduction plans to build the model. The data came from three urban areas in the United States: Los Angeles, Tucson, and Washington State that have had trip reduction requirements on employers for many years. Employers were required to submit plans to reach a particular objective such as a reduction in the levels of single occupant vehicle (SOV) use. The data consisted of worksite modal characteristics aggregated at the employer level and a listing of incentives and amenities offered by employers. The Los Angeles data contained the largest data set data, with the Tucson, Arizona and Washington state data sets being considerably smaller. Data quality control problems reduced the size of the data in each area and eliminated or restricted some potentially useful variables (e.g., dollar values of some incentives). For performance evaluation the datasets were divided in two disjoint sets 'training/testing set' which was used to build the models and 'validation set' which was used as an unseen data to evaluated the models.

The dependent variable chosen was the change in vehicle trip rate (VTR) (e.g., reduction of 4.5 vehicles per 100 employees). VTR correlates closely with the goals of TDM -- reduce trips, decrease air pollution, decrease the need for parking -- and is generally proportional to the desired result. Alternative dependent variables such as SOV share or average vehicle ridership (AVR) have disadvantages. SOV share misses the benefits of moving from one non-SOV mode to another where the switch may actual reducing traffic but not affect SOV share (e.g., carpool to transit). The reduction in vehicle trips is distorted when using AVR as the dependent variable due to the non-liner relationship between AVR and vehicle trips.

For example, increasing AVR by 0.25 from 1.10 to 1.35 persons per vehicle would require a reduction in 17 vehicle trips per 100 people. The same increase (0.25) for a worksite with an AVR of 1.50 to 1.75 would only require a reduction in 9 vehicle trips per 100 people.

Two approaches were used for the model building process: linear statistical regression models and non-linear neural networks. The linear statistical regression models were used as a benchmark for the validity and accuracy of the neural net models. The linear statistical regression models minimize the sum of the error between the real and predicted data, learning simple linear relationships between the worksite characteristics, incentives and the dependant variable ‘change in VTR’, while the neural networks learn more complex non-linear relationships. Sometimes linear regression methods were used to determine which variables the neural net would use to build its models.

Several phases were followed to build the models. Models were built for each of the three datasets using a variety of approaches of handling the data, including variable selection, grouping of incentives, and the treatment of outliers. Models were also built after combining the data from the three urban areas into a single dataset. Under the assumption the transportation industry was most interested in a model that predicted when large reductions could be achieved, the model performance objective was focused on predicting the change equally well across the range of the changes in VTR.

Los Angeles

Phase I

The Los Angeles dataset consisted of 25,459 total records. A first attempt to build neural network models using all the variables present in the data resulted in very poor performance. So, different linear regression modeling approaches were used to select different variables sets on which neural network models were built. The neural network model built using stepwise method variable set was able to get better results on all performance measures.

Phase II

Next, the Los Angeles data was over-sampled in some bins and under-sampled in others in order to improve the accuracy on moderate range of change in VTR. In addition, this data was added with the costs associated with each incentive, and a stepwise neural net model was built. It was concluded that the over-sampling biased the model towards predicting more negative changes in VTR with reduction in overall accuracy.

Phase III

Later it was discovered that some of the records in the data sets had not a single incentives implemented. It was felt that any worksite which goes from having some incentives in one year to no incentive the next year and again having some incentive the next to next year, was missing the information regarding the incentive plans for the intermediate year. So it was decided that these records be removed as they might influence the model in some wrong direction. The cleaned dataset now contained 18,140 records in total. Also, it was felt that a simple model which was not significantly worse than the best complex model would be more preferred due to its simplicity. So all of the individual incentives were replaced by the grouped incentives and news models were built. The two best models obtained from this model-building session were the ones constructed using the neural network: one built using stepwise regression variables on incentives and costs and the other one built on grouped incentives. Both models got 16.92% accuracy on full range of change in VTR, with comparable accuracies on moderate range of change in VTR and R-square values.

Phase IV

Finally, it was felt that by removing the records with very large changes in VTR (which might not be the effect of incentives offered) might improve models performance. So a normal distribution of the data set was produced and all records which were outside of three distributions above or below the mean were removed from the set. In effect 790 records were dropped from the validation set and 7,432 from the training set. Three models were built with variables based on previously constructed models. However, the best model, the one built from stepwise regression on incentives with costs added was still lacking in accuracies when compared to the models from previous phases.

Recommended Los Angeles Model

The two candidate models were the ones built on data with records with 'no incentives' removed using the stepwise regression variables on incentives and costs and the other built on grouped incentives. There was no significant difference between their accuracy measures of the two models. So the simple model built on data with grouped incentives was regarded as the recommended Los Angeles Model.

Tucson

Phase I

The Tucson data set consisted of 1,121 total records. The models built with variable selection did not select many of the incentives variables in predicting change in VTR and so were considered unsuitable. The neural network model built with no variable selection got the best 'bin classification on full range of change in VTR' of 20.54%, 16.60% accuracy on moderate range of change in VTR and the R-square value of 0.022.

To reduce the complexity of the previous models, all of the individual incentives were replaced with grouped incentives. The neural network model built with no variable selection seemed to over-fit the training data and got poor results on validation data. While the linear regression model was better than the neural network model it was still worse than the previously built neural network model on individual ungrouped incentives

Phase II

In this phase, the Tucson data was over-sampled to boost the accuracy in the moderate range of change in VTR. A neural network model built with no variable selection which got 18.75% accuracy on full range of change in VTR, 12.77% accuracy on moderate range of change in VTR and an R-square value of 0.036 gave balanced predictions.

A better model on over-sampled grouped incentive data was the neural network model having 16.96% accuracy on full range of change in VTR, 14.86% accuracy on moderate range of change in VTR with 0.007 R-square.

Recommended Tucson Model

In the end, two neural network models with no variable selection were front runners for the best model: one was built on full sample data with individual incentives, and the other was built on the over-sampled grouped incentives data. The winner was evidently the neural network model built on individual incentives on full sample data.

Washington

Phase I

The Washington data set consisted of 1,414 total records. Since some variables in the data might not be readily available to employers or program coordinators who might be using the software, employee's preferences towards incentives and the type of work variables were removed from the data. In the study it was found that the employee's preferences towards incentive variables did indeed influence the change in VTR. However, since these variables may not be available to future software users, they were only used to study their impact. The two models which came out as front runners were the linear forced enter regression model and the neural network model both built on all variables except for preferences.

To reduce the complexity of the models, all of the individual incentives were replaced with grouped incentives and simple neural network with no variable selection and a linear forced enter regression model were built. The neural network model built showed a better distribution of accuracy in the moderate range of change in VTR than the linear regression forced enter model. The scatter plots also showed that the predictions of the neural network model built with all grouped variables were much closer to the actual values.

Phase II

Since better accuracies were desired on the moderate range of change in VTR, an over-sampling of the Washington data was done. Again the models with employee's preferences towards incentives came out better restating their importance in predicting changes in VTR. But, since the preference variable information might not readily accessible to everyone, the models built without preference variables were considered for final analysis. The forced enter regression models from this model-building session came up to be the best, but still it performed worse than its equivalent model on original full sample data.

Two more models were built with grouped incentives over-sampled data. The better model was the neural network built with no variable selection as it seemed to be less biased than the linear regression model.

Recommended Washington Model

As stated in the previous sections, the neural network models built in both phases I and II data including employee's preferences towards incentives obtained much better accuracies than the models built without them. However, given that this information might not be available to future model builders, only models without preferences were considered.

The two models that were considered as the best on the Washington data were the forced enter linear regression model built on the individual incentive full sample data and the neural network model built on grouped incentive over-sampled data. The models were so close that a cross-tab of positive/negative actual/predicted changes in VTR was set-up to decide. The most important factor of the best model was its ability to predict correct negative changes in VTR. So it was desirable to improve recall which captured the completeness of the model at predicting negative changes in VTR without sacrificing precision which captured the correctness of the model at predicting negative changes in VTR. The f-value weighted these two measures equally (with $\beta = 1$) to give one performance measure. The forced enter linear regression had the best f-value on validation set so was considered the recommended model for the Washington data set.

Combined Data Models

All of the incentive data from each of the three data sets was combined to build a larger, generalized model, based on the information of all three areas. Some of variables had to be collapsed into one group variable in order to ensure uniformity of the incentives.

Phase I

The training data from all three cities was combined into a single training set, while the validation sets for each of the cities were left intact for testing of the combination training sets. Neural net with no variables selected models and forced enter linear regression models were built on data containing only 23 variables. The models were evaluated on the three validation sets and then compared with best independent model from three locations. It was found that the neural net model built on the combined data with no variable selection was able to get better 'bin classification accuracy on moderate range of change in VTR' on the Los Angeles validation set as compared to the recommended independent Los Angeles model built on grouped data. The combined data neural net model also was able to improve bin classification for the recommended independent Tucson model built on over-sampled grouped incentive data. The Washington data set and recommended model did not have improved accuracy with the combination neural net model.

It was concluded from this neural network model with no variable selection that by adding more data from the other datasets to Los Angeles data helped in improving the accuracy of the Los Angeles model, and therefore became the recommended model for the Los Angeles data. However, the Tucson data, while increasing accuracy on moderate range of change in VTR at expense of accuracy on full range, became more biased to predicting negative VTR changes and the Washington data actually showed a decrease in accuracy with the Phase I combination neural net model.

Phase II

To deal with the combined data neural network model's problems with the Washington and Tucson data sets, the three training sets of data were sampled equally to get equally combined training data set. This data set contains 2,018 records from each data set. A neural network and a forced enter linear regression model were built on this new equally sampled data. The equally sampled combined data set models still could not improve the accuracy on the Washington validation set. The equally combined data neural network model was able to improve accuracy on Tucson validation set on moderate range of change in VTR with reduced R-square value. But still it was difficult to state that it was the best model for Tucson data. So a cross-tab was done, and the neural network model built on the equally sampled combined data resulted in better Recall F-value on the Tucson validation set and therefore became the recommended model for the Tucson data.

Best Generalized Model

Overall, the best generalized model for any location is the neural net model built on equally sampled data based on the three performance measures described earlier.

Table 63 Performance Measures for Best Generalized Model

	LA Validation	Washington Validation	Tucson Validation	Training Set
Bin Classification Accuracy on moderate range of change in VTR (i.e. bins a2 to a5)	23.20%	14.62%	25.32%	29.24%
Bin Classification Accuracy on full range of change in VTR (i.e. all bins)'	15.78%	12.50%	20.54%	21.09%
R-square	0.075	0.011	0.013	0.215

This is the model at http://www.nctr.usf.edu/worksite.

VII. Conclusion

No single variable selection technique, data handling method, or modeling approach yielded the best-fitting model for all urban areas. In many cases, there was no significant performance difference between the top models, so the recommended model for some dataset had to be decided by using the F-value measure, which incorporated two other metrics: *Recall* which gave a measure of the completeness of the model, and *Precision* which gave a measure of the correctness of the model.

The best model for each city also was not the model that used data only from that city. Before combining the Los Angeles data set with those from the other two areas, the preferred model was the one built on the grouped incentives data with records with 'no incentives' removed. But after combining the datasets, the neural network model built with no variable selection performed better for Los Angeles than the model built with only data from Los Angeles. Also for Tucson data, a neural network model built on the equally sampled (i.e., each dataset contributes equally) data performed better than the previously selected neural network model built on the full sample (i.e., all valid records from three datasets) with ungrouped incentives data. The best model for the Washington data was the linear forced enter regression model built on full sample with ungrouped incentive data.

The generalized models for any urban area were built on the combined training datasets and equally sampled training datasets. The models built using equally sampled datasets were the ones which were not biased towards the any dataset. So the best generalized model for any location is the neural net model built on equally sampled data which is the version deployed at http://www.nctr.usf.edu/worksite.

Overall, the neural net models performed better than the linear regression models. This might be due to the ability of the neural network program to move beyond simple linear regression, which tries to minimize the error between the predicted and actual data sets. The neural network models in many situations were able to learn the non-linear relationships among various combinations of strategies. There were some neural network models which performed worse than the linear models. This might be due to the over-fitting of the training data and reducing the neural net's power to generalize over unseen validation data.

Quality control issues with the provided datasets affected the model building process. In the case of the Los Angeles data, there were many worksites for which some incentives were available in one year, then not a single incentive was shown in the following year, only to have incentives "reappear" the next plan cycle. These unexplained gaps in reporting can affect the ability of the model to estimate the impact of a particular incentive.

Another data problem encountered was the use of different units of measurement across programs. For example, the transit subsidy values were reported ranging from $0.20 to $3,000. Though employers were to report these values as “cost per employee per month”, the lower figure basis appears to be along the lines of “cost per employee per trip” and the upper figure might be the “total cost per employee using the mode per year”. The difficulty is there is no way of telling from the data. Also it was found that the incentives had differences in their definition across different datasets which introduced error into the results. This problem made model-building a complex task when trying to condense and collapse all of the similar variables into one.

The aggregate nature of the data loses the ability to explain whether the change in mode behavior was influenced by the programs or changes in the workforce or other exogenous variables. While hundreds of thousands of employee data, including employee’s preferences to particular options, were available from the State of Washington dataset, there was no identifier to track individual changes in behavior over time. The other two datasets did not have any comparable data to Washington’s individual survey responses. Access to such disaggregate data could improve the ability of a model to track behavior changes over time based on changes to worksite incentives, amenities and programs. As suggested in the Future Work section, the development of a nationwide database which would include employee’s preferences for incentives, similar to those used in the State of Washington data.

VIII. FUTURE WORK

Given the common interest shared by the public sector and worksites in assessing the relative effectiveness of worksite trip reduction program, future work should begin by improving the quality of the data already being collected. Quality can be improved by adopting standard definitions and common terminology. Common terms will contribute to an expanded dataset by making the data compatible with other data from other parts of the country. This approach could be facilitated by the creation of a centralized database.

Adhering to quality control procedures also could add more explanatory power to the data, especially as it relates to the value of financial incentives. Improving the quality of information that already exists can help worksites more cost-effectively deliver vehicle trip reduction programs.

The data used in this model-building approach was aggregated to the worksite level prior by the employer. This aggregate level detail does not allow for analysis to determine which individuals' travel behavior has changed. In order to get a real handle on what makes VTR increase or decrease and what causes people to choose alternative transportation options, attempts to control for the differences should be used. Access to disaggregate data collected over time can help establish a "test" and "control" group approach to control for differences, for example, in the composition of the workforce.

Access to the disaggregate data also would help track the long-term effects of the programs. The current project assesses the impacts between two time periods (usually separated by only one to two years). However, the cumulative effect of these programs over time is less understood (i.e., will the worksite experience a constant, variable, or exponential change in VTR over time as the programs diffuse within the workforce and move beyond the "early adopters"?) *Diffusion* is the process by which the trip reduction program offerings are communicated through certain channels over time among the employees at that location. While the collection of individual data may be difficult due to privacy and attrition issues, it is worth investigating the possibility of collecting this type of transportation behavior data to help develop sustainable transportation strategies and programs for the future.